Memory for Processing Sequence in Cognitive Skills and Its Role in Performance Errors

Final Technical Report
for the Period 1 May 1997 through 31 July 2000
for AFOSR Grant F49620-97-1-0219 (Cognition Program)
Dan J. Woltz and Michael K. Gardner
University of Utah

Submitted to:
US Air Force Office of Scientific Research
Air Force Systems Command, USAF

AFOSR/NL
801 N. Randolph St., Rm. 732
Arlington, VA 22203-1977

REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-01-

0051

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
15-01-2000	Final Technical Report	01-05-1997 to 31-07-2000

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Memory for Processing Sequence in Cognitive Skills and Its Role in Performance Errors	F49620-97-1-0219

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Dan J. Woltz and Michael K. Gardner	

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
University of Utah Salt Lake City, UT 84112	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR/NL 801 N. Randolph Street, Room 732 Arlington, VA 22203-1977	AFOSR/NL
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A

Unlimited   Approved for Public Release
Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
We conducted a series of laboratory experiments to develop a theoretical understanding of how sequential cognitive skills are represented in memory, how memory for processing sequence facilitates skill performance under some performance conditions, and how it degrades skill performance under other performance conditions. We found that memory for processing order plays a substantial role in performance improvements that result from consistent practice, however it can also lead to performance errors in near transfer tasks ("strong-but-wrong" errors). Individual differences issues were also studied.

15. SUBJECT TERMS

Cognition, human error, skilled performance, human memory

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
Unclas	Unclas	Unclas		69	19b. TELEPHONE NUMBER *(Include area code)*

Summary

We conducted a series of laboratory experiments to develop a theoretical understanding of how sequential cognitive skills are represented in memory, how memory for processing sequence facilitates skill performance under some performance conditions, and how it degrades skill performance under other performance conditions. We examined these issues in a variety of skill tasks that varied in both content and complexity.

We found that memory for processing order plays a substantial role in the performance improvements that result from consistent practice in sequential cognitive skills. We also found that after substantial amounts of practice, memory for processing order can be triggered inappropriately by relatively minor changes in task demands (i.e., near transfer trials that demands familiar processing operations but in a slightly different order). When learners are exposed to multiple processing sequences during training, they are typically unaware of the stimulus changes that induce these errors. The near transfer errors are most frequent when a processing sequence begins like a familiar one, but ends differently. The likelihood of these near transfer errors increases with practice on a sequential skill. Furthermore, the vast majority of these errors are undetected by the performer without explicit feedback. In total, these errors closely resemble "strong-but-wrong slips" that have been hypothesized in describing some performance errors observed in real-world performance environments. The current laboratory demonstrations are important because they verify the existence of and delineate the underlying mechanisms responsible for this error type that had previously been proposed only in post hoc analysis of real-world errors.

In studies investigating the nature of sequence knowledge that underlies strong-but-wrong errors, two conclusions were supported. First, memory for processing sequence appears to be implicit rather than explicit in nature. That is, skill performers who demonstrate the acquisition of sequence knowledge in patterns of performance facilitation, are unable to recall or recognize the sequences that they have learned. Second, sequence knowledge appears to be represented as a chain of linked associations, rather than as a set of unitized components that execute in an all-or-nothing fashion. The latter, rejected form of representation corresponds to mechanisms of composition and chunking that have been proposed in some theories of skill acquisition.

In studies investigating individual differences in error-proneness, our evidence supported several conclusions. First, some individuals consistently make more strong-but-wrong, near transfer errors than other individuals following extensive skill practice. Although working memory capacity predicts differences in early phases of skill performance, it did not predict near transfer errors. Neither did other conventional cognitive abilities (e.g., processing speed and knowledge). However, measures of attention disengagement ability accounted for a large share of variance in transfer errors. This suggested that error proneness at late stages of skill acquisition depended in large part on the inability to control attention processes rather than lack of general cognitive resources.

Finally, we investigated several training factors that could reduce strong-but-wrong near transfer errors. Evidence strongly supported a role of sequence variability during training in reducing these errors. A training method that attempted to induce

flexibility in attention focus did not reduce near transfer errors, although it did reduce performance latency.

In total, this research extends current theoretical understanding on the nature of skill acquisition and representation for an important class of cognitive skills: Those that require a sequence of linked processing steps. The research has also resulted in a theoretical understanding of the mechanisms underlying a type of error that could be consequential in real-world performance environments. Although highly trained individuals in any complex skill seldom make major errors, reports of performance slips that have disastrous outcomes regularly occur in military, medical, and industrial settings. The current work advanced our understanding of how undetected slips are induced in highly skilled performers by environmental conditions that closely resemble conditions encountered during training.

Research Objectives

A long-standing finding in the research literature is that higher levels of skill in cognitive tasks result in faster and more accurate performance (e.g., Bryan & Harter, 1899; Crossman, 1959; Fitts & Posner, 1967; LaBerge, 1973; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). While this trend holds true in most circumstances, recent theoretical advances in cognitive science (e.g., Anderson, 1983, 1993) lead to the prediction that experts (i.e., highly skilled performers) will be more error prone under certain training/transfer circumstances than novices. Furthermore, theory predicted that these errors should be unavailable to conscious introspection -- that is, they should be undetected by the performer. Coinciding with this theoretical prediction, applied research on human error in real-world job environments suggested that highly skilled individuals exhibit 'mental slips' under some circumstances, and that these slips can be disastrous if undetected.

The research reported here represents a series of empirical studies aimed at testing these predictions and understanding the mechanisms underlying such errors in skilled performance. Earlier research (Woltz, Bell, Kyllonen, & Bell, 1996) suggested that a common category of undetected slips might occur after individuals are highly practiced in sequential skills (i.e., skills that require a specific sequence of operations). Furthermore, the evidence suggested that many of these errors could be the result of inappropriate execution of strong memory for familiar sequences. The primary purpose of the research reported here was to develop an understanding of how sequence memory acquired from the practice of sequential skills is represented in memory, how and when this memory results in performance errors, what individual characteristics are associated with error proneness, and whether different forms of initial skill training might reduce the likelihood of slips.

Background

When individuals practice skills that require a specific sequence of cognitive operations, performance can benefit from several forms of learning. Some performance

improvements probably derive from increased familiarity with the overall task structure. Other improvements can come from the strengthening of the component operations, if they were not learned to asymptotic levels prior to their use in the sequential skill (Schneider, 1985). In some skills, performance improvement can develop from the representation and retrieval of entire problem instances that have been extensively practiced (Logan, 1988). Finally, individuals can benefit from acquiring knowledge about common sequences of operations, if they are consistent over problems (e.g., Carlson & Lundy, 1992; Woltz, Bell, Kyllonen, & Gardner, 1996). Evidence suggests that sequence knowledge is distinct from both component and instance knowledge. This paper addresses questions concerning this element of skill acquisition in sequential, multi-step cognitive skills.

Evidence for the positive impact of processing sequence knowledge has come from research on several types of perceptual-motor and cognitive skill tasks. Nissen and others have demonstrated the positive effect of participants acquiring sequence knowledge in a simple perceptual-motor task, serial reaction time (Cohen, Ivry, & Keele, 1990; Curran & Keele, 1993; Nissen, Willingham, & Hartman, 1989; Willingham, Nissen, & Bullemer, 1989). Lewicki (Lewicki, Czyzewska, & Hoffman, 1987; Lewicki, Hill, & Bizot, 1988) and Stadler (1989) demonstrated a role of sequence knowledge in a primarily perceptual processing task. The acquisition of artificial grammars might also be considered a form of sequence learning (e.g., Mathews et al., 1989). Finally, processing sequence effects have been studied in a variety of computational tasks. Elio (1986) demonstrated the importance of one form of sequence knowledge in a multi-step computational task that had participants calculating numeric indexes from given values (also see Frensch, 1991). Charness and Campbell (1988) demonstrated a form of sequence knowledge when participants practiced a multi-step algorithm for squaring 2-digit numbers. Carlson and his colleagues have demonstrated the role of sequence knowledge in other numeric computation tasks similar to that of Elio (Carlson & Lundy, 1992; Lundy, Wenger, Schmidt, & Carlson, 1994; Wenger & Carlson, 1996) and in binary computation sequences such as logic gates (Carlson & Shin, 1996; Carlson, Sullivan, & Schneider, 1989; Carlson & Yaure, 1990). Finally, in earlier work we demonstrated the unique role of learning sequences in a novel computation skill referred to as number reduction (Woltz, et al., 1996).

It is unlikely that the same type of memory representation is responsible for all of these demonstrations of sequence learning. In some cases, the sequence knowledge is partly motor (e.g., Fendrich, Healy, & Bourne, 1991) and in other cases it apparently has little or no motor component (e.g., Cohen et al., 1990; Stadler, 1989). Sometimes the nature of the sequence knowledge may depend on the consistency of data to be operated on within the sequences (Carlson & Lundy, 1992), and sometimes it probably does not (e.g., Woltz et al., 1996). Most of the time the acquisition of sequence knowledge appears to be independent of reliable declarative knowledge about the processing sequences (Cohen et al., 1990; Knowlton & Squire, 1996; Lewicki et al., 1987; Mathews et al., 1989; Nissen & Bullemer, 1987; Nissen et al., 1989; Willingham et al., 1989). However, some studies have found an association between implicit and explicit measures of sequence information (e.g., Fendrich et al., 1991; Lundy et al., 1994). In addition, some evidence suggests that sequence knowledge may be acquired differently, depending on attentional demands of the learning environment (Curran & Keele, 1993).

Given the different task paradigms and varying evidence about processing sequence knowledge, it is not surprising to find different theoretical explanations of these learning effects. Some researchers have concluded that sequence knowledge reflects a restructuring process during skill acquisition such as production composition, chunking, or step-skipping (Blessing & Anderson, 1996; Charness & Campbell, 1988; Frensch, 1991; Frensch, 1994; Lundy et al., 1994). These explanations suggest that practice of a consistent set of sequential operations results in a new memory representation of the original operations, typically one that represents the sequence as a whole unit. In contrast, other researchers have suggested that sequence knowledge may be represented independently from the operations used in the sequence (MacKay, 1982, 1987). In MacKay's hierarchical network theory, independent representations of sequence knowledge produce anticipatory priming in the component operations. Present evidence does not conclusively rule out any of these theoretical positions, and it is possible that some restructuring mechanisms such as composition occur under some conditions but not others (e.g., Carlson & Lundy, 1992).

Despite the diversity of evidence regarding the nature of processing sequence knowledge in sequential cognitive skills, several general conclusions seem warranted. First, the impact of sequence knowledge is evident in a variety of tasks with considerably different processing demands. Second, in some sequential skills the acquisition of sequence knowledge has a greater impact on performance than the learning of other skill components (e.g., Charness & Campbell, 1988; Woltz et al., 1996). Third, regardless of the effect size, sequence knowledge has usually enhanced rather than degraded overall skill performance. Even when the introduction of new sequences during transfer has produced performance decrements in new sequences relative to old ones, typically there has been positive rather than negative transfer overall (i.e., performance on new sequences was still better than if there had been little or no practice on old sequences). However, we will describe some evidence that suggests negative transfer can result from the acquisition of strong processing sequence knowledge.

Negative Transfer Errors in Cognitive Skills

Convincing demonstrations of negative transfer have been relatively infrequent in the skill learning literature. Singley and Anderson (1989) argued that while personal anecdotes of negative transfer are common (e.g., people complain of interference from using a new version of a computer software program or from driving in Great Britain after learning to drive in the United States), convincing experimental evidence is rare. Most likely, individuals who experience such interference in a familiar skill are still more efficient than those who had little or no prior practice, particularly when considering the aggregate of many task components. In total, there may be positive transfer under the new learning conditions, although performance might be impaired somewhat in comparison to performance in the more familiar conditions. Furthermore, the impairment may be limited to a few task elements that are compensated for by efficiency in other elements. Transfer is considered negative only when skill performance under new task conditions is worse for those individuals with more training than for those with less training.

The most widely cited demonstrations of negative transfer in skill performance and problem solving are the water jug experiments by Luchins (1942). Participants in

these studies solved a series of water jug problems, where the object was to measure a specified quantity of water using various combinations of three jugs with known volumes. The problems were arranged such that one algorithm worked as a solution to the first five problems (e.g., Jug B - Jug A - 2 x Jug C). Subsequently, problems were of two types. Some could be solved by the familiar algorithm, as well as by a simpler algorithm. A second type of transfer problem could only be solved by a new algorithm and not by the familiar one. Several experiments demonstrated the inflexibility produced by practice on the consistent algorithm problems. Practiced individuals favored the familiar but complex solution over the simpler solution when either would work. More importantly, practiced participants performed more poorly than unpracticed participants on problems that could not be solved using the familiar algorithm. In one study using college students, only 39% of the practiced participants could find a solution to this type of problem, whereas all of the unpracticed participants obtained the correct solutions.

The set effects found by Luchins appeared in a sequential cognitive task. The training problems required a consistent sequence of operators on data that differed from problem to problem. However, it is unclear whether participants were simply learning a response (i.e., the response 1B – 1A – 2C or a verbal equivalent), or a sequence of operations. If participants had merely memorized a response, then the negative transfer would reflect the use of an inappropriate shortcut strategy rather than procedural sequence knowledge.

If the water jug task and its analogs were the only skills for which negative transfer errors could be demonstrated, then the importance of the phenomenon could be questioned, especially given the plausible strategy explanation. However, in earlier work we found evidence of negative transfer in a different cognitive skill (Woltz et al., 1996). The number reduction task was a variant of a computation skill developed by Thurstone and Thurstone (1941) for assessing individuals' ability to learn mental procedures. In its simplest form, the skill has two rules for reducing multi-digit stimuli to single digits. The stimulus numbers can contain any combination of the digits 1, 2, and 3. When two contiguous digits are the same, they can be reduced to a single digit of that value (e.g., 22 = 2, 11 = 1, and 33 = 3). When two contiguous digits are different, they can be reduced to the remaining digit (e.g., 23 = 1, 31 = 2, 12 = 3, etc.). In multi-step problems that have more than two digit stimuli, participants apply the reduction rules in sequence from left to right, carrying forward a temporary solution from each operation to the next. For example, the stimulus 132 first requires the application of the different rule (13 = 2). The answer (2) is combined with the next digit (2), which then requires the application of the same rule (22 = 2).

Unlike the water jug problems, a single response cannot be learned. Different problems that require the same sequence of operations typically require different responses. For example, 132 and 213 both require a different-same rule sequence but yield different final answers, 2 and 3 respectively. In addition, participants were exposed to a mixture of problems that reflected more than one sequence, so simple solution strategies were less likely. The results indicated that sequence knowledge had a greater impact than instance knowledge on performance latency. Of primary importance here, negative transfer errors were evident in one experiment. Participants with more training on a subset of the sequences made more errors compared to those with less initial training when exposed to new sequences.

Even considering this number reduction evidence, clear experimental evidence of negative transfer in skill performance is rare, and the generality of the phenomena must be questioned. To address this, we consider non-experimental evidence that may represent negative transfer in natural settings. Several attempts have been made in the past few decades to study errors people make in familiar settings of everyday life and in various work environments. We examined these taxonomies of everyday errors in relation to the existing evidence for negative transfer in sequential or multi-step cognitive skills.

Taxonomic Categories of Action Slips

Three recent attempts have been made to classify everyday human errors with respect to their underlying cognitive and conative processes (Heckhausen & Beckman, 1990; Norman, 1981; Reason, 1990). These three taxonomic theories have emerged from the analysis of highly similar and partly overlapping databases of recorded action slips. Interestingly, each taxonomy divides this largely common set of slips in different ways. One clusters the slips by the nature of underlying memory representations and mechanisms (Norman, 1981), one by temporal sequence of underlying intentional processes (Heckhausen & Beckman, 1990), and one by modes of performance reflecting attention and knowledge (Reason, 1990). As such, each taxonomy paints a somewhat different overall picture of error mechanisms. However, despite their differences, one pervasive theme across taxonomies is that many errors can be described as the inappropriate influence of strong habits related to, but distinct from, the intended action. In all three theories, this phenomenon is described when performance of some task is relatively automatic, and often when an environmental distraction is present.

Norman (1981) described these errors as capture slips. Capture errors occur when a strong habitual action sequence is substituted for a related, weaker action sequence. For example, an intention to drive to a new store may result in driving to a more familiar store, especially if the routes partially overlap and if some distractions are present (e.g., an engaging conversation with a passenger). Norman attributed these errors to the faulty activation of child schemata (i.e., activating an action plan for going to a familiar store) in order to satisfy the goal of a parent schema (i.e., going shopping).

Heckhausen and Beckmann (1990) proposed a similar error category termed sidetracking errors. As with Norman's capture errors, sidetracking errors were attributed to a strong habitual action substituting itself for some other intended act that overlaps in processing components. Heckhausen and Beckmann suggested that this substitution is most likely when individuals rely on automated behavior to achieve high level goals. They refer to this as a wide rather than narrow goal span of attention.

Finally, Reason (1990) describes the same phenomenon within a category of slips termed strong-but-wrong errors. These errors are presumed to occur in both rule-based (i.e., procedural) performance and skill-based (i.e., automatic) performance. Reason linked the predominance of strong-but-wrong errors to similarity matching and frequency gambling tendencies, which are thought to be pervasive in the cognitive system. Reason (1990) summarizes these principles with respect to errors in stating, "when cognitive operations are underspecified, they tend to default to contextually appropriate, high-frequency responses" (p. 97).

These error categories within the three taxonomies are conceptually equivalent to negative transfer errors. When a strong but inappropriate procedure intrudes on an action sequence, it implies that prior experience or training is responsible for the error. Individuals with stronger tendencies (i.e., more experience or training) would be the most susceptible to these errors. Thus, although there is limited experimental evidence of negative transfer in skill performance, the error taxonomies suggest that it may reveal itself regularly under certain real world conditions.

The detection of errors by a skilled performer is also an important issue, both practically and theoretically. A slip by an air traffic controller or physician is more likely to lead to disaster if it goes undetected and therefore uncorrected. Moreover, errors that are immediately detected by the performer appear to be qualitatively different from those that are undetected for several reasons. They suggest a different level of attention or self-regulation involved in the skill performance. Undetected slips seem more likely when performance is based on implicit rather than explicit memory processes and when the focus of attention is at a more global level (i.e., when attention is not required for local operations that are highly familiar).

All three taxonomic theories propose that error detection rests on the monitoring and comparison of intentions to actions, where actions are almost exclusively depicted as overt motor responses. According to this perspective, undetected errors occur when no monitoring of the actions is taking place (e.g., attention is captured by something else in the environment), or when the intention is specified at a different level from the action (e.g., there is a global intention such as driving home from the work place, and the automatic actions to accomplish this are specified at a lower, data-driven level). Although purely cognitive skills, especially those that require an ordered sequence of operations, are not directly addressed by the theorists, we presume that slips in these skills would be considered undetectable because many of the actions are mental transformations, and as such their products cannot be compared to intentions.

Overview of Methodological Issues in the Current Research

Conducting laboratory experiments to investigate theoretical questions concerning skill performance errors presents unique methodological challenges that we have been forced to address. Prior to presenting our primary experiments, we describe these methodological issues and how we have addressed them in our research.

Most laboratory researchers face the tradeoff between controls in the experimental setting that are ideal for addressing specific research questions and the ecological validity of the experiments in representing realistic real world conditions. Investigations of skill performance errors in the laboratory may represent an extreme example of this tradeoff. There are three interrelated reasons for this claim.

First, attempting to investigate performance errors made by individuals who are highly practiced in some set of cognitive operations is in itself a difficult endeavor. Obviously, in learning experiments that provide extensive practice in a skill task, participants' overall performance improves and the likelihood of most mistakes decreases. Thus, by the nature of our topic, we are investigating relatively low frequency events. This creates rather severe problems of reliability of the dependent measures. It is

common in cognitive research to use response latency from correct responses as the dependent measure. With this measure, the researcher typically can estimate with some precision the number of observations needed per individual to obtain adequate reliability. However, when errors are the focus, often only a fraction of the trials represent errors. Consequently, more observations are usually needed per person. Furthermore, it is often difficult to estimate the number of total trials needed to obtain reliable measures given individual differences, and this is especially difficult if the response latency for error responses is of interest.

Second, the type of error that we have chosen to study was theorized to occur only among skilled individuals. So, in virtually all of our experiments we had to provide enough practice such that participants were approaching asymptotic levels of performance in the skill. For practical reasons, this demanded that we investigate relatively simple skills that could be mastered in a few experimental sessions. This, of course, limits the generalizability of our findings to some extent in relation to very complex real-world skills. While we have had to design our experiments under this general constraint, we have attempted to investigate the degree to which our findings generalized across different skills and different degrees of complexity within skills.

Third, investigating errors that primarily represent mental slips undetected by the performer requires special attention to how participants perceive the nature of the experiment. If participants know that the researcher is interested in errors, then they adopt an attitude of unusual carefulness to monitor for and reduce errors. Similarly, performance feedback can alter the way participants approach the experimental task. Certain experimental manipulations produce marked increases in error rates in some skill tasks. Participants are typically unaware of the number of these errors, unless accuracy feedback is provided. If made aware of the high error rate, most subjects will dramatically alter their performance strategy. However, without any feedback during skill acquisition, most participants experience frustration.

After extensive pilot testing of different combinations of task instructions and feedback, we established a standard method for conducting most of our experiments. Typically, participants would perform between three and five training sessions, depending on the complexity of the skill task. During training, accuracy feedback was provided after each error response so participants could correct inaccurate declarative or procedural knowledge about the skill. In addition, after each block of training trials (about 30 trials in most experiments), summary error and latency feedback was provided. Furthermore, during the training sessions, we set performance goals of responding as quickly as possible while making no more than 10% errors. The summary feedback allowed participants to monitor how well they were achieving these goals. Corrective feedback was provided if a participant's performance deviated too far from the performance goals. If participants made too many errors (e.g., 15% or more), the computer advised them to slow down to be more careful. If participants made no errors, the computer advised them to try and respond more quickly even if they made a few errors.

In the typical experiment, the training sessions were followed by a transfer session that contained our manipulations needed to test various research questions. However, because we wanted to investigate mental slips associated with subtle changes in the task demands, we attempted to conceal the task manipulations. Toward this end,

each transfer session typically began with a number of training trial blocks that closely resembled those presented in previous training sessions. The only difference was that no accuracy feedback was provided after either trials or blocks. The performance goal was also changed slightly, and participants were instructed to respond quickly while making no errors. In experiments in which we investigated error awareness, we also gave participants the opportunity to retake trials in which they thought they had made an error. This allowed us to distinguish between errors that were detected and corrected and those that were undetected. After several blocks of training trials using these instructions, new transfer trials were intermixed with the familiar training trials. Most hypotheses about sequence memory and errors were tested by contrasting performance on the new and old trials in this phase of the transfer session. Participants, however, were not informed of the insertion of new trials, and different forms of evidence suggested that most were unaware of the task manipulations. In total, we believe that these methods allowed us to investigate near transfer errors that were often undetected by the performer in skills that had been learned to asymptotic levels of performance.

Basic Demonstration of Sequence Knowledge and Transfer Errors in Simple and Complex Sequential Skills

In this set of two experiments, we investigated the potential cost of the acquisition of strong processing sequence knowledge in a multi-step cognitive skill. Prior research had primarily demonstrated the performance benefits of sequence knowledge in a variety of perceptual-motor and cognitive skills. We tested the link between sequence knowledge and a form of negative transfer that has been supported by a considerable amount of anecdotal evidence, but little experimental evidence. These experiments are reported in greater detail in Woltz, Gardner, & Bell (2000).

In the first experiment of this series, we used the simple, 2-rule version of number reduction described earlier. Here we tested the existence of negative transfer reported by Woltz, et al. (1996) using longer sequences. In addition, we tested whether negative transfer errors reflected the relatively automatic application of strong-but-wrong sequences or the application of slower attention driven processes following the recognition of new task demands. Finally, we contrasted two types of sequences. One sequence type allowed the learning of specific response patterns, similar to Luchins' water jug problems. The second type allowed only the learning of processing sequences that were not tied to specific response patterns. We contrasted these to determine whether einstellung-like errors are dependent on consistent response patterns.

In Experiments 2, we used a more complex version of number reduction that required participants to learn many more sequences that combined four rather than two rules. The purpose here was to assess whether negative transfer errors that corresponded to the einstellung and strong-but-wrong descriptions could be found in a sequential skill that better represents the complexity of many real world cognitive skills. In Experiment 2, we assessed the existence of negative transfer errors (i.e., whether more practice led to more errors), whether these errors reflected strong-but-wrong procedural memory or

novel problem solving mechanisms, and the extent to which these errors were detectable by the participant.

Experiment 1. In this experiment, we used a simple version of the number reduction skill that had been used in previous research on memory for processing sequence (Woltz et al., 1996). In the previous research, we had participants solve 3-digit problems that required only a 2-rule sequence. In the current experiment, participants reduced 4-digit problems that required 3-rule sequences.

Processing sequence refers to the order of three component rules needed to solve any 4-digit stimuli. For example, 3213 would be first reduced to 113 by applying the different rule to the first two digits (i.e., 32=1). Then 113 would be reduced to 13 by applying the same rule (i.e., 11=1). Finally, 13 would be reduced to 2 by applying the different rule. In this example, the stimulus 3213 can be solved only by applying the Different-Same-Different rule sequence. In this first experiment, 198 participants learned the skill by solving problems representing two distinct sequences that balanced the frequency of each component rule at each of the three positions. Each sequence was represented by 6 different instances of the sequence (e.g., 1232, 1323, 2131, 2313, 3121, and 3212 were all instances of the Different-Same-Different sequence). Two experimental groups practiced 12 such instances of the number reduction task. The groups differed only with respect to the amount of practice that they had prior to transfer trials. Both groups had the same amount of instruction regarding the component rules, but one group had fives times as much practice in applying the rules in sequence.

The transfer trials included all of the previously practiced instances of the two training sequences (old sequence trials), and instances of two new sequences not seen in training that used the same component rules with equal frequency at each position. New transfer sequences were created such that each one matched an old training sequence in the first two component rules, but ended with a different rule. Negative transfer would be demonstrated if high skill participants made more errors than the lower skill participants on new sequence trials. Thus, we expected that the partial match of stimulus conditions to strong memory for processing sequences would result in executing incorrect operations in the final problem step.

We also predicted that if the high skill participants made more errors on the new sequence transfer trials, the response latency for these errors would be fast, similar to the latency for correct responses to old sequence trials. Given that we could produce errors in the new sequence transfer trials, this prediction was important to discriminate between weak-method explanations (Anderson, 1989) and skilled memory explanations for these errors.

Finally, we contrasted two subsets of the sequences to examine whether negative transfer errors were associated with response patterns. Luchins' evidence from water jug experiments could be explained by memory for a single response rather than memory for a sequence of operations. Therefore, it was important to determine whether negative transfer was dependent on this feature, or whether negative transfer in sequential skills can result from abstract memory for order of operations.

Figure 1 presents the mean latency and error data by problem step for both high and low skill conditions. Blocks 1-20 represent the training phase, and Blocks 21-30 represent transfer.
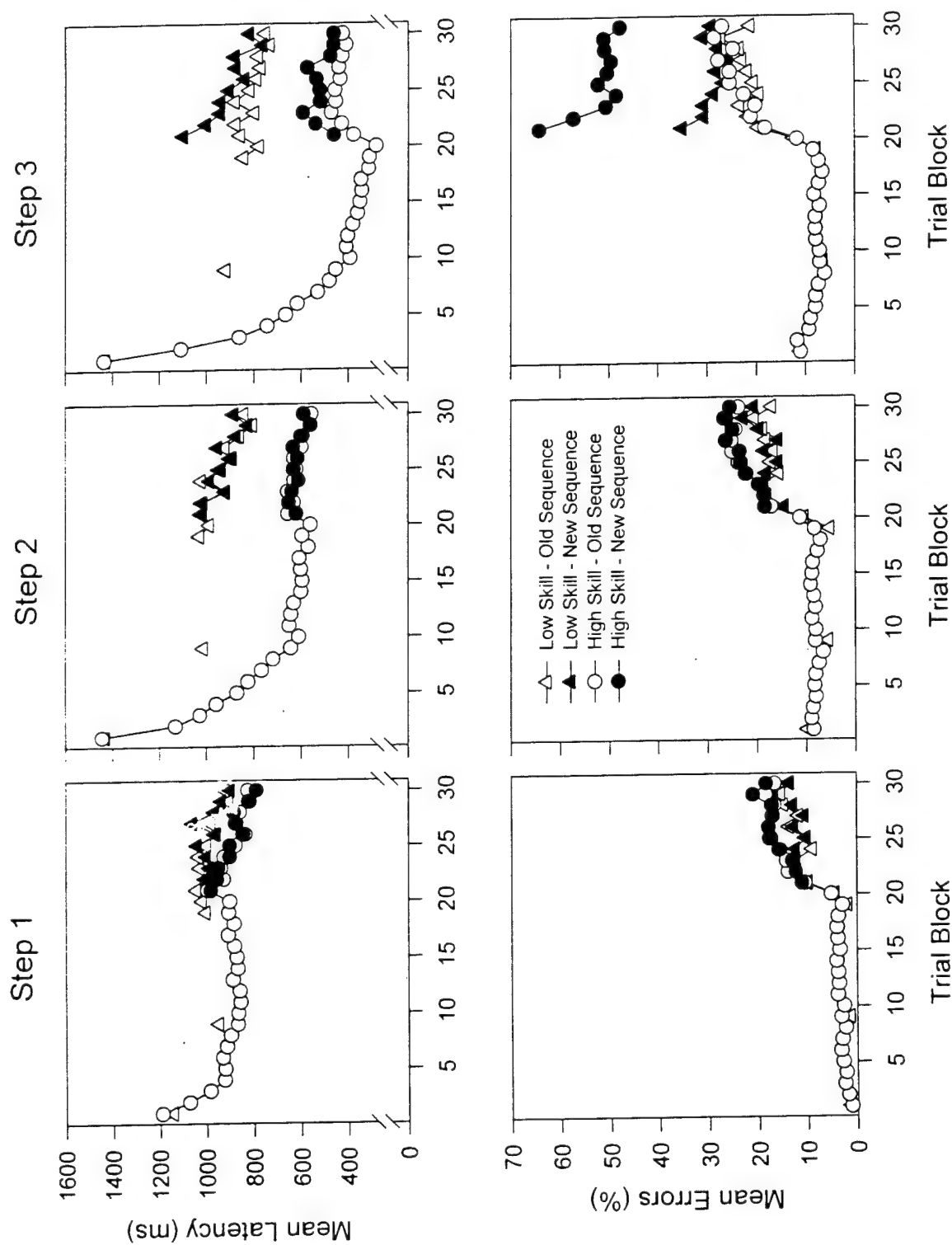
Figure 1: Mean Latency and Error Data for Experiment 1 by Skill Level and Trial Condition

The hypothesized errors due to sequence memory were tested in Step 3 of the transfer blocks, where new sequence trials differed from old sequence (training) trials. The data supported the hypotheses in that error rates differed between old and new sequence trials exclusively in Step 3, p < .05, and that this old-new effect was greater for high skill participants, p < .05 (see Blocks 21-30 in the lower panels of Figure 1). This finding is consistent with the prediction that more practice in sequential cognitive skills leads to negative transfer when new transfer sequences resemble familiar ones in the initial steps.

Two other features of the transfer error data were important. First, although the high skill group showed the greatest old-new error difference, the low skill group showed a significant old-new effect. Even after just four blocks of practice on the old sequences (96 trials), the low skill group made more errors on new compared to old sequence trials, p < .05. However, the high skill group made considerably more errors than the low skill group initially, and they continued to show a large old-new error difference through all 10 blocks of transfer. So, the low skill group showed transfer errors on new sequences, but the errors were considerably lower in magnitude and less persistent than those seen among high skill participants.

Second, the introduction of new sequence trials disrupted old sequence performance. This effect can be seen most clearly by comparing the last two blocks of training trials (Blocks 19 and 20) with the old trials in the first two blocks of transfer (Blocks 21 and 22). These blocks were presented contiguously without any instruction to the participants that trial content would change. This disrupting effect appeared in both groups and in all three steps. Across all three steps, the low skill group old sequence error rate changed from $\underline{M}$ = 7.05 % ($\underline{SD}$ = 7.60) to $\underline{M}$ = 16.64% ($\underline{SD}$ = 18.62), and the high skill group error rate changed from $\underline{M}$ = 7.95% ($\underline{SD}$ = 7.31) to $\underline{M}$ = 16.56% ($\underline{SD}$ = 17.90). The overall effect on old sequence performance of adding new sequence trials was reliable, p < .05, but the group difference in this effect was not, p > .10.

Although our primary research questions pertained to performance errors during transfer, we analyzed the latency data to better understand the processes likely to underlie the errors. Of primary importance was the latency on correct versus error responses on new sequence trials. If error responses were slow relative to correct responses, then errors most likely reflected failures during deliberate retrieval and application of the rules when unfamiliar sequences were encountered. If error responses were as fast or faster than correct responses, then errors more likely reflected strong-but-wrong application of skilled memory representations (i.e., some form of procedural memory for the sequence of operations). This inference would be further supported if latency on the new sequence errors were equivalent to latency on correct old sequence trials, which would presumably rely on the same skilled memory representations.

As described earlier, the upper panels of Figure 1 show the mean latency data by trial block for correct and incorrect responses combined. As can be seen here, transfer latency differed between new and old sequences only in Step 3. Figure 2 presents the mean latency data only for Step 3, broken down by correct and incorrect responses to old and new transfer trials. In an effort to obtain more stable latency data, we included only participants who made at least two errors in each trial condition. Because of this, the means in Figure 2 were computed on a subset of the total sample (Low Skill n = 65; High
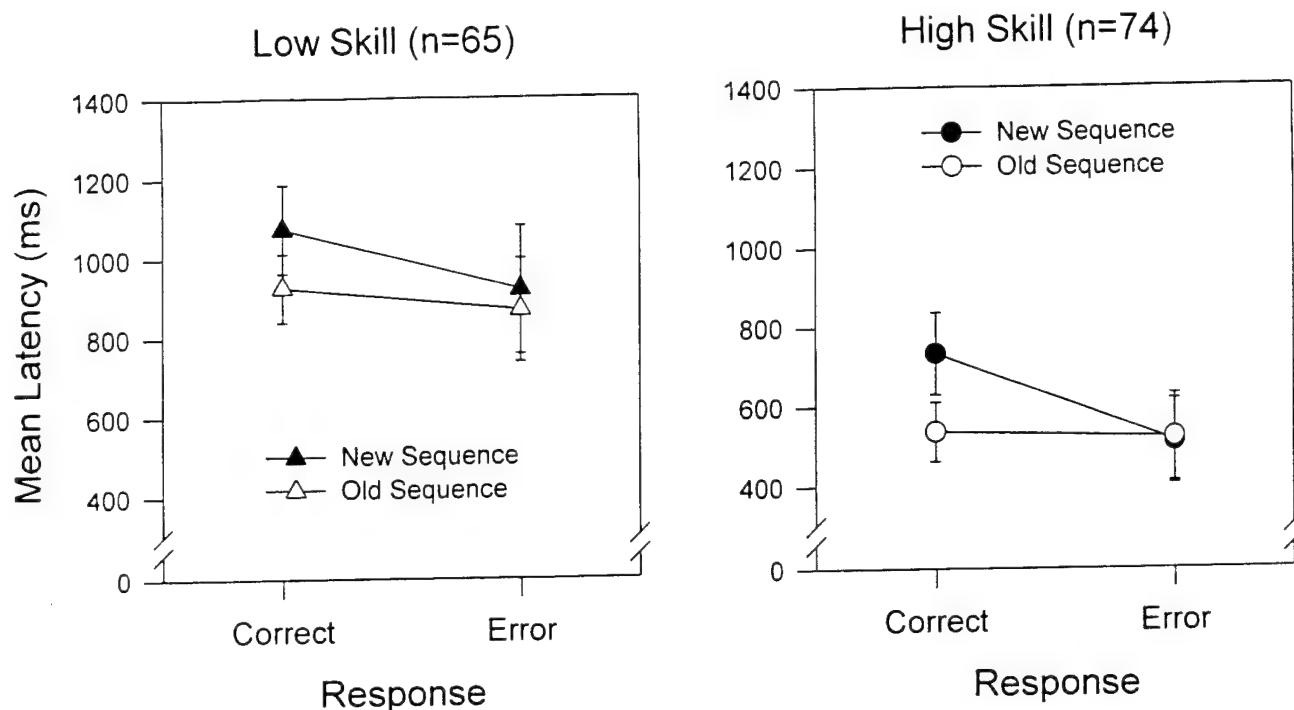
Figure 2: Mean Latency for Transfer Trials by Response Type and Trial Type

Skill n = 74). As seen in this figure, mean response latency for old trials was equivalent regardless of whether it was a correct or incorrect response. However, for new sequence trials, correct responses took longer than incorrect ones. This pattern was statistically significant, p < .05. These data are consistent with the interpretation that errors on new sequence trials in the third step represented the misfiring of a strong-but-wrong procedural memory for the old sequence rule.

A final question addressed in this experiment pertained to the distinction between learning response patterns versus rule patterns. As described earlier, the original einstellung demonstrations in the water jug problems may have depended on the acquisition of a response (e.g., B – A – 2C, where A, B, and C were water jugs with varying volumes). This would be a form of step-skipping described by Blessing and Anderson (1996). In the number reduction task, learners could acquire partial knowledge of response sequences on half of the training sequences, thus allowing for step-skipping.

Figure 3 presents mean errors on Step 3 by trial type and skill level for each sequence in the experiment. Each participant either received DSD and SDS as training sequences or DSS and SDD. Whichever sequences were not seen in training became the transfer sequences. If the transfer errors seen in Figure 1 were due primarily to response anticipation, then new sequence errors in Figure 3 should have occurred only in two of the sequences. Specifically, when the same rule was last in a training sequence (DSS or SDS), participants could have learned to anticipate the final step response (i.e., simply a repeat of the Step 2 response), and this would result in many errors in the corresponding transfer new sequence (DSD or SDD). However, when DSD or SDD was a training
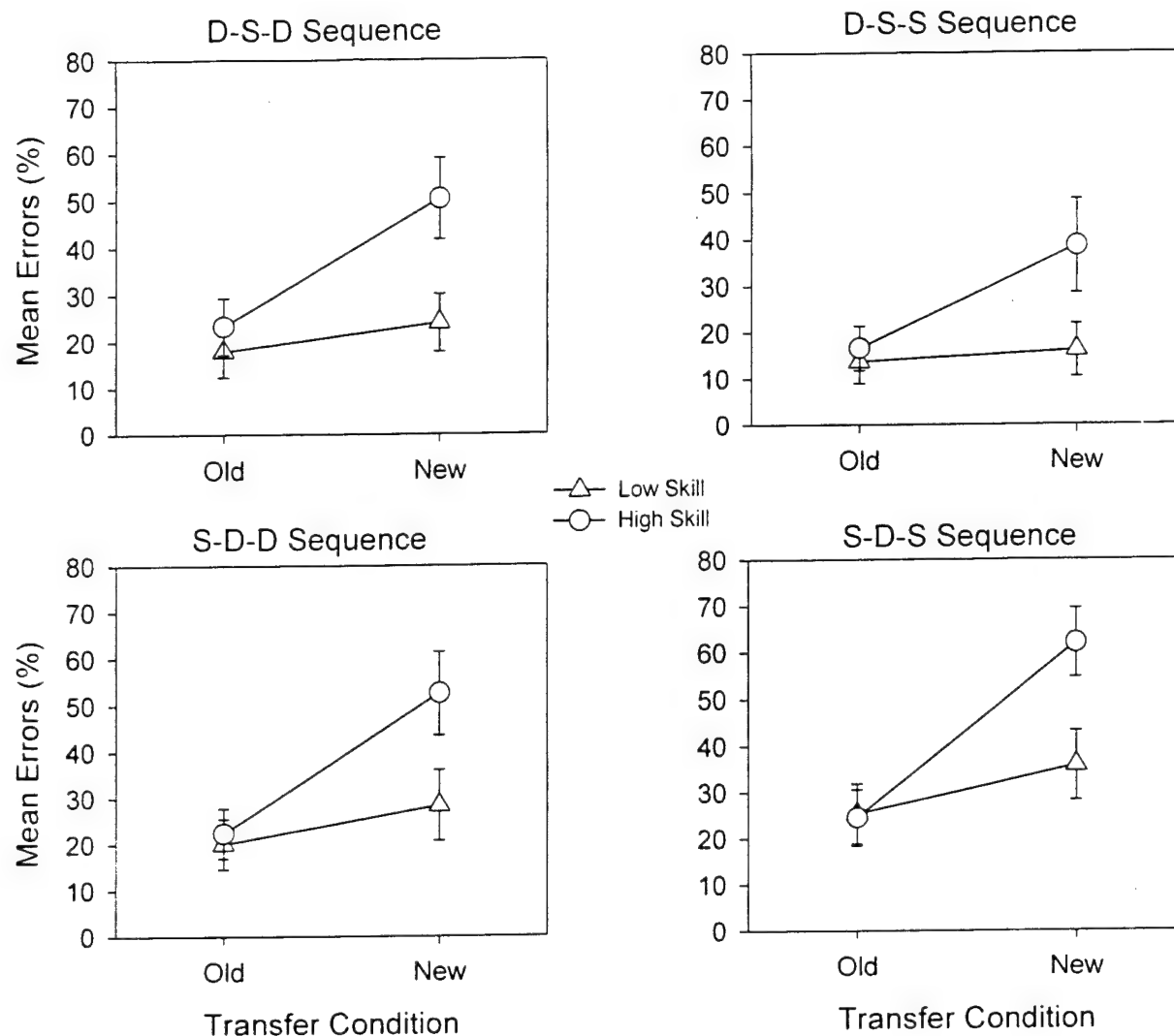
Figure 3: Mean Errors by Skill Level, Transfer Condition, and Sequence

sequence, no such response anticipation was possible. Thus, if the transfer errors were primarily due to response anticipation, we would expect few new sequence errors on the corresponding transfer sequences (DSS and SDS).

As seen in Figure 3, all four sequences exhibited a substantial old-new error difference for high skill participants. This outcome is inconsistent with a strict response anticipation explanation of the transfer errors. Instead, it suggests that the strong-but-wrong errors were more likely due to sequence learning that manifests as rule anticipation.

Experiment 2. This experiment had a similar structure to Experiment 1 of this set, but we investigated three additional questions. First, we tested whether the negative transfer errors observed in Experiment 1 could be produced within a more complex skill that involved practice on many processing sequences typical of real-world cognitive skills. Second, we tested whether the errors were attributable to sequence knowledge or instance knowledge. Third, we assessed the degree to which transfer errors were detectable by participants.

We used a complex version of number reduction that also had been used in previous research (Woltz et al., 1996). In the current version of this task, stimuli could consist of any combination of the digits 1 – 9. Reduction was accomplished by applying some combination of four component rules. The same rule, which was identical to the simple version, states that two identical numbers could be reduced to a single digit of that same number (e.g., 77=7). The midpoint rule states that two numbers that differed by two could be reduced to their midpoint (e.g., 53=4). The contiguous rule states that two numbers in either an ascending or descending sequence could be reduced to the next number in the sequence (e.g., 32=1; 67=8). Finally, the last rule states that two numbers whose difference is greater than two could be reduced to the last of the two numbers (e.g., 28=8; 63=3). As before, these rules were applied to multi-digit stimuli by parsing the stimuli pair-wise left to right and carrying forward intermediate solutions to be combined with the next digit in the stimulus. For example, 9687 would be first reduced to 687 by applying the last rule to the first two digits (i.e., 96=6). Then 687 would be reduced to 77 by applying the midpoint rule (i.e., 68=7). Finally, 77 would be reduced to 7 by applying the same rule. Unlike the simple version of number reduction, participants entered only the final response to each problem rather than all component responses[1].

As in the simple version of Number Reduction, processing sequence refers to the order of three component rules needed to solve any 4-digit stimuli. In the previous example, the stimulus 9687 can be solved only by applying the Last-Midpoint-Same rule sequence. In contrast to the first experiment in which participants were trained on two sequences with six instances per sequence, participants in this study were trained on 12 different sequences, each with 12 instances (e.g., 9687, 1534, 3798, and 8312 are all instances of the Last-Midpoint-Same sequence). As before, the set of sequences balanced the frequency of each component rule at each of the three positions. Because the task was more complex, we provided more training. As in Experiment 1, two experimental groups differed only with respect to the amount of practice that they received prior to a transfer condition. The low skill group had one session of practice, and the high skill group had four sessions.

The transfer condition included all of the previously practiced instances of the 12 training sequences (old sequence-old instance trials), new instances of the 12 training sequences (old sequence-new instance trials), and 12 new sequence trials not seen in training. As before, new sequence trials were created such that each one matched an old training sequence in the first two component rules, but ended with a different rule. Thus, we expected that the partial match of stimulus conditions to strong memories for processing sequences would result in increased errors. If transfer errors were attributable to abstract sequence memory rather than instance-specific memory, then the errors should be restricted to new sequence trials and not new instance trials. As before, negative

transfer would be demonstrated if the high skill group showed the predicted new sequence errors more than the low skill group.

We expected the negative transfer errors to be primarily undetected rather than detected errors. This reflects our presumption that the errors are due to relatively automatic execution of procedural knowledge, which is inaccessible to conscious awareness. Accordingly, we attempted to discriminate between detected and undetected errors during transfer trial performance. We did this by instructing participants to correct all errors by pressing a key that allowed them to retake any trial. This, coupled with instructions and feedback during transfer that emphasized the importance of error-free performance, allowed us to separate errors that were recognized by the performer from those that were not.

Finally, we predicted that if the high skill participants made more undetected errors on the new sequence transfer trials, that the response latency for these errors would be fast, similar to the latency for correct responses to old sequence trials. In the event that we produced errors in the new sequence transfer trials, this prediction was important to discriminate between effortful problem solving (i.e., deliberate attempts to apply declarative rules) and skilled memory explanations for these errors.

A total of 72 subjects participated in this experiment (n=34 high skill and n=38 low skill). Mean data from the training sessions are summarized in Figure 4. Blocks 1-10 represent the initial number reduction training session for both groups. These blocks occurred in the first experimental session for the high skill group and in the fourth experimental session of the low skill group who performed an unrelated skill task for the first three sessions. Blocks 11-55 were training blocks performed only by the high skill group in their second, third and fourth experimental sessions. Blocks 55-60 were the initial training blocks of the fifth session for both groups. These blocks differed from the others shown in Figure 4 only in that instructions to make no errors were introduced. As seen in this figure, participants conformed reasonably well to the instructions of maintaining an accuracy rate of approximately 90% in the first four training sessions. Furthermore, extended practice resulted in systematic reductions in response latency.

Figure 5 presents the mean latency data for the transfer blocks of the final session by trial condition and group (these immediately followed Blocks 55-60 shown in Figure 4). The 10 transfer blocks were collapsed into five 2-block sets for analyses. We analyzed these data to test whether the high skill participants would show greater degrees of instance-specific and sequence-specific performance facilitation.

First, the data revealed an overall difference between the high and low skill groups, $p < .05$. Consistent with what was seen in the Session 5 training blocks, the high skill participants were approximately 500 ms faster than the low skill participants across all transfer trial conditions. More importantly, the effect of trial type depended on skill level. As can be seen in Figure 5, high skill participants showed a greater difference between new and old sequence trials than did low skill participants (i.e., high skill participants showed more sequence-specific facilitation), $p < .05$. This interaction was consistent with the prediction that more practiced participants would show greater evidence of sequence learning.
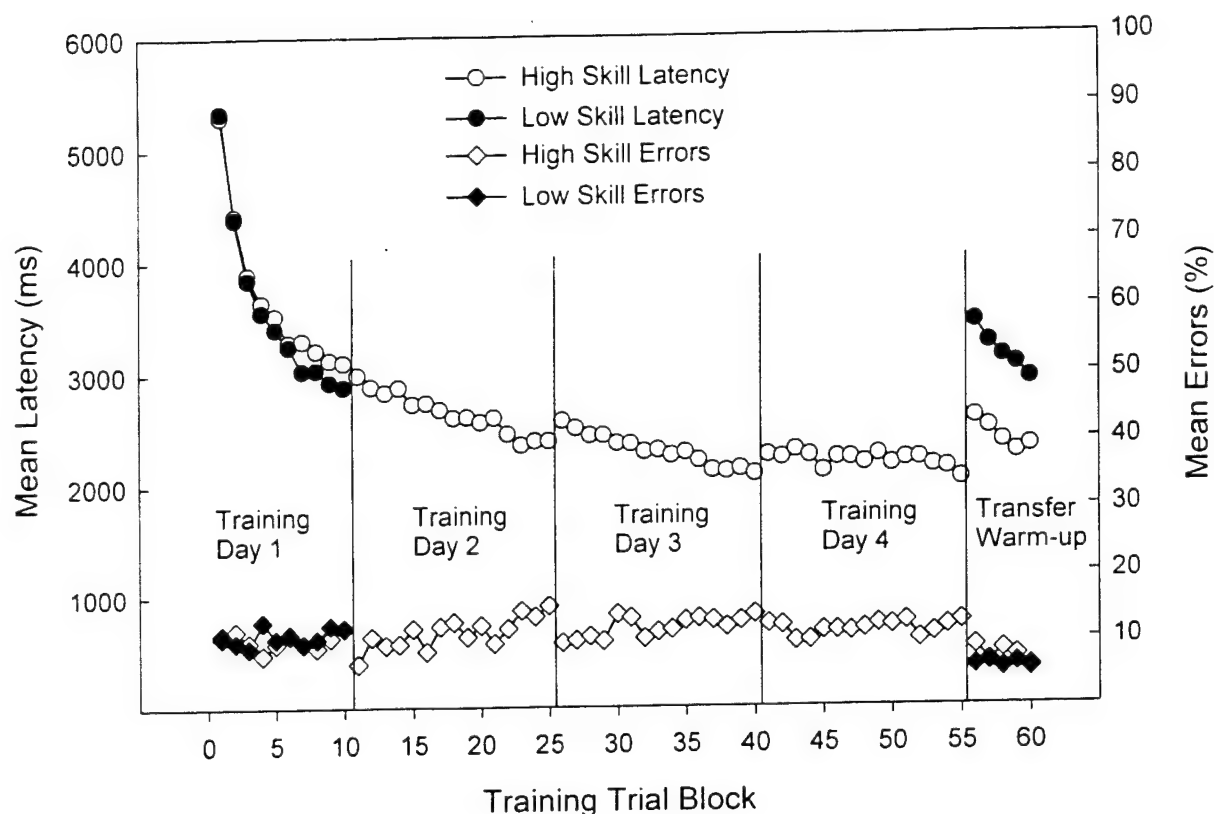
Figure 4: Mean Latency and Error Data for Training Trials in Experiment 2

In contrast to the difference in ɔequence-specific facilitation between the two skill level groups, there was no difference in the degree of instance-specific facilitation. The overall effect of old versus new instances was significant, p < .05, but the Group x Old-New Instance interaction was not significant, p > .10. Thus, the high skill participants differed from the low skill participants in the degree of sequence memory effects, but not in the degree of instance memory effects.

While the latency data lent support to our basic assumptions about the effects of practice on the performance of this sequential processing skill, the main hypotheses of this study were tested with the error data. Figure 6 shows the mean error rates for both high and low skill participants in the various transfer trial conditions. In addition, errors were broken down into detected and undetected categories. Detected errors represent incorrect responses that were corrected by the participant by pressing the spacebar and retaking the trial. Undetected errors represent incorrect responses that were not followed by a spacebar response.
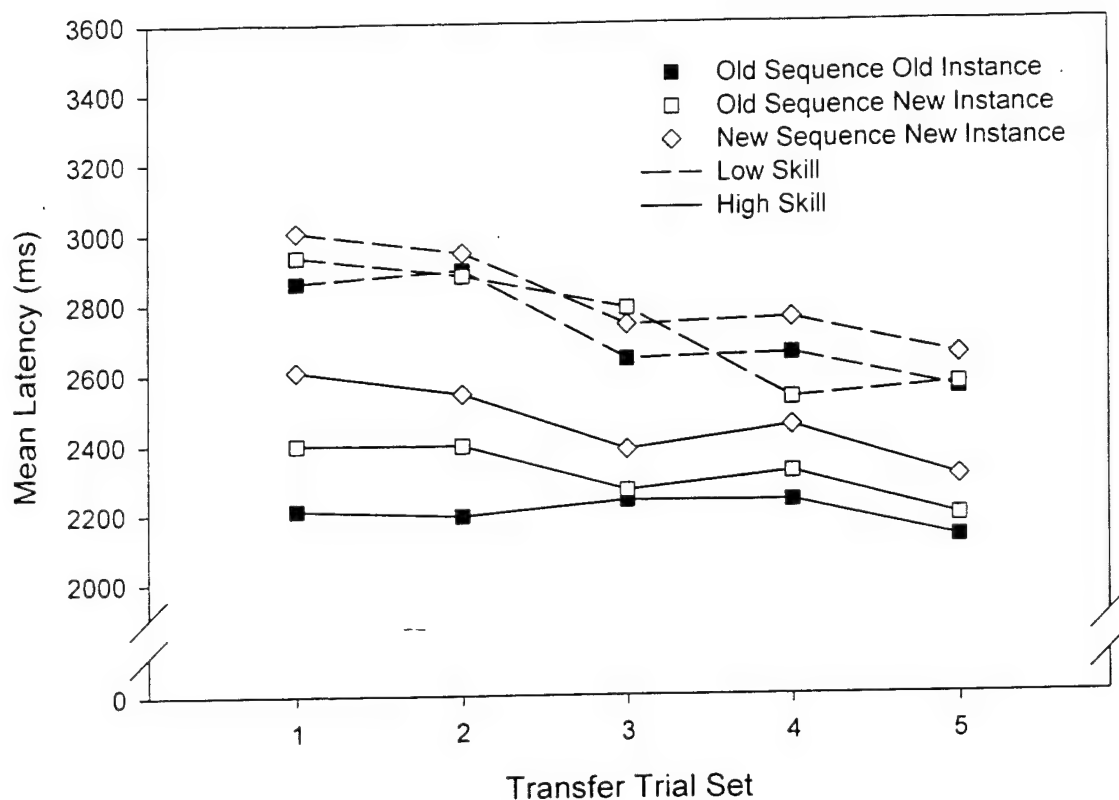
Figure 5: Mean Latency for Transfer Trials of Experiment 2 by Trial Type

As seen in Figure 6, the high and low skill participants were equivalent in their error rates for all categories of trials and errors except one, $p > .10$. As hypothesized, the high skill participants made substantially more errors on new sequence trials compared to old sequence trials, $p < .05$, and the additional errors were almost exclusively undetected. The high skill group made nearly twice as many undetected errors on new sequence trials as did the low skill group.

In total, the error data are consistent with our hypothesis that high levels of cognitive skill, which involve the representation of processing sequence information for familiar trial types, will lead to more undetected errors when new processing sequences that resemble the old ones are introduced. Such errors resemble the einstellung errors demonstrated by Luchins (1942), but they occurred within a skill that required the learning of 12 rather than one sequence. However, from the error data alone it was not clear that these errors represented the fast execution of strong-but-wrong procedures. They could have reflected slower, more effortful processing associated with recognizing new task demands. To distinguish these interpretations, we examined the latency data on the undetected errors. If undetected errors on new sequence trials represented the misfiring of skilled memory representations for old trials, the latency of these errors

should have been similar to that for correct old trials. If undetected errors on new sequence trials represented inaccurate weak-method processes, then their latency should have been considerably slower than that for correct old trials.
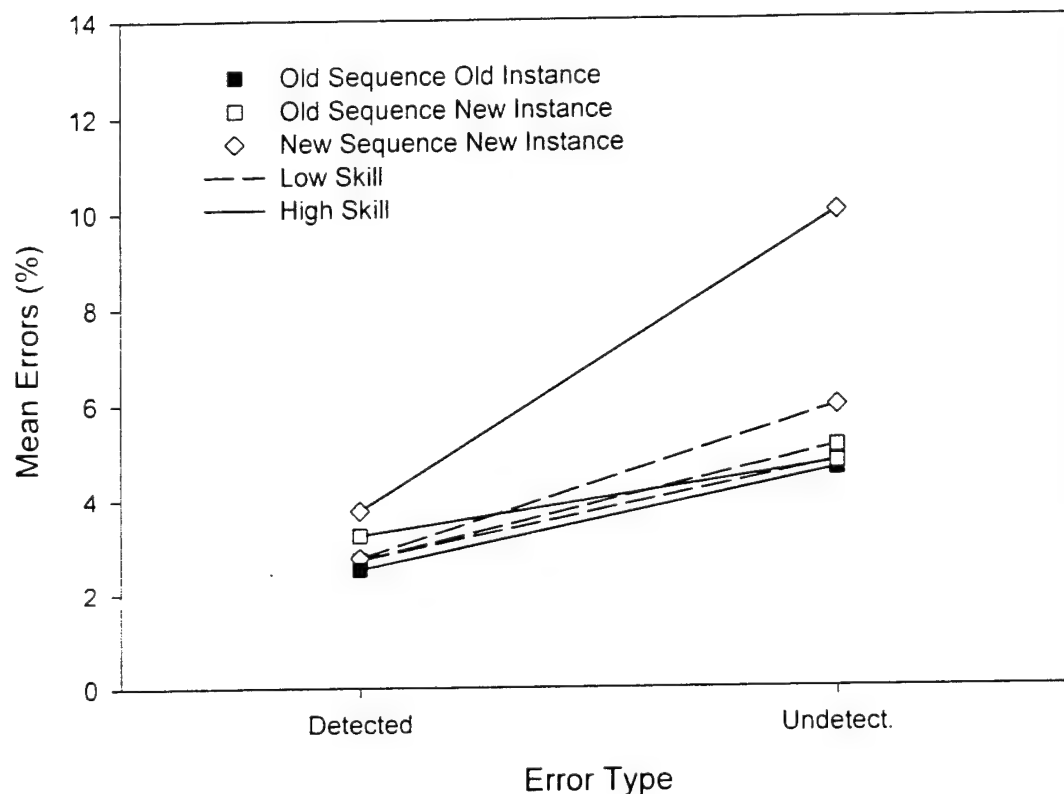

Figure 6: Mean Transfer Errors by Trial Type and Skill Level

Figure 7 presents the median latency values for correct responses and for undetected error responses for those participants who made two or more undetected errors on both new and old sequence trials (n=29 for the low skill group and n=24 for the high skill group). There was virtually no difference in these latency values for the two types of old sequence trials (new and old instances), so these categories were collapsed because it allowed more participants to meet the criterion for inclusion in the analysis.

As seen in Figure 7, the latency data for undetected errors generally supported our skilled memory hypothesis. That is, the average latency for undetected errors on new sequence trials for high skill participants did not differ reliably from the average latency for correct responses on old sequence trials, p > .10. In contrast, the latency for undetected errors on new sequences was reliably less than the latency for undetected errors on old sequences, p < .05. Thus, for the high skill group, these data supported the interpretation that undetected errors on new sequence trials may have relied on the same

procedural memory processes as correct responses on old sequence trials. An unexpected result seen in Figure 7 for the high skill participants was the longer latency on undetected errors on old sequence trials. Because participants made relatively few undetected errors on old sequence trials, we viewed this effect somewhat skeptically without further replication.
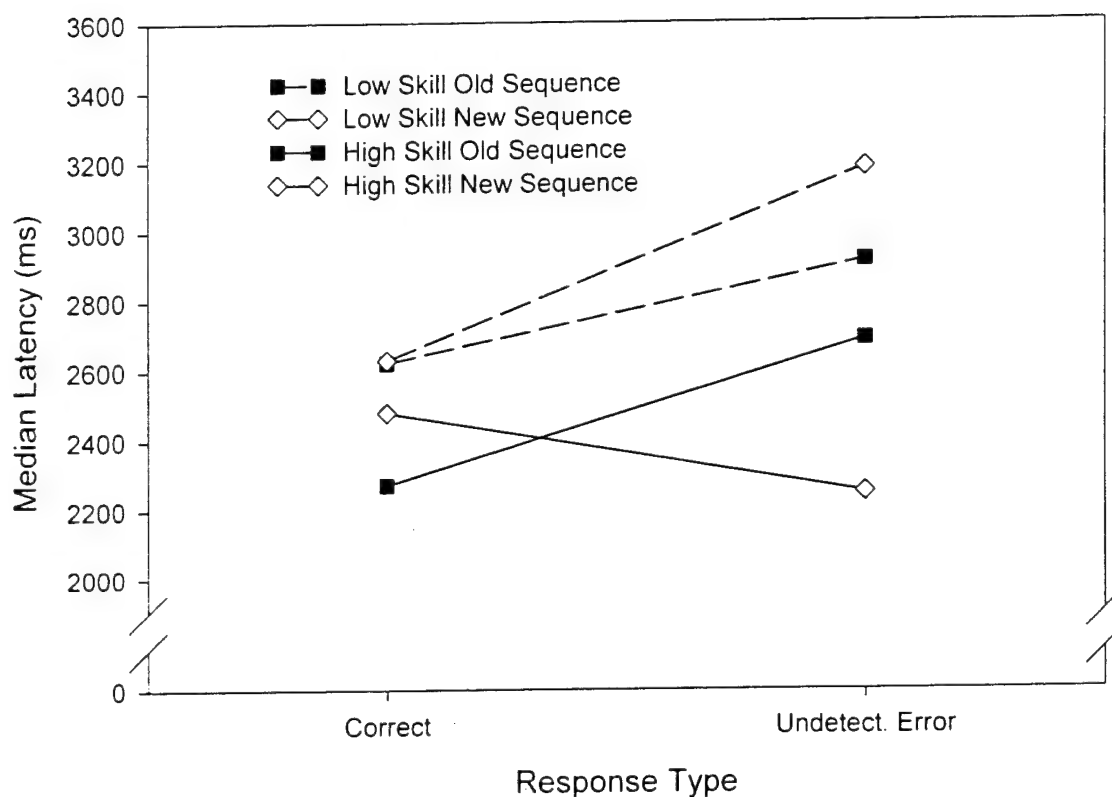

Figure 7: Median Latency for Transfer Trials by Response Type, Sequence Type and Skill Level

The latency data shown in Figure 7 for low skill participants revealed a different pattern than that found for the high skill participants. Here the latency for undetected errors on new sequence trials was longer than latency for correct trials on old sequences, $p < .05$, and it did not differ from latency for undetected errors on old sequences, $p > .10$. Thus, only the high skill participants made fast, undetected errors on new sequence trials. Low skill participants, who did not show evidence of sequence learning in the earlier latency analyses, also did not show evidence that their undetected errors on new sequence trials were driven by the same processes as correct responses on old sequence trials.

Verbal Protocol Analysis of Error Types in a Complex Skill

By examining think-aloud protocols from subjects performing a cognitive skill task after extensive practice, we found evidence supporting our earlier interpretations that new processing sequences differing only slightly from familiar sequences induce strong-but-wrong undetected errors. In critical trial components, more than 60% of the errors were accompanied by verbal responses that indicated the firing of a wrong (but familiar) operation for that component. This, combined with our earlier evidence, represents the only laboratory evidence that we know of for this phenomenon in skilled performance.

In one study, we had participants performing five sessions of the complex number reduction skill. In the first four sessions, each subject is exposed to 12 3-rule sequences. In the fifth session, an additional 12 new sequences are introduced. Each new sequence matches an old sequence on the first two rules. Previous evidence from this task showed that the new transfer sequences induced undetected errors. In the current study, we video taped participants during performance, and asked them verbalize each intermediate answer and the final answer, but to still try to perform as quickly as possible. As in previous study, subjects could retake any trial for which they thought they made an error by pressing the space bar. This allowed us to separate detected from undetected errors.

For each error during the transfer session, we coded information from both the video segment and data recorded by the computer with respect to verbal intermediate responses, possible subject comments, time intervals between verbalized answers, whether the error was detected (trial retaken), and unusual events taking place on the prior trial. We then attempted to categorize each error according to the rule in the sequence in which the error occurred (first, second, or third), and whether the error was a misapplication of the appropriate rule, whether a wrong rule was applied, or whether the rule was appropriate and applied correctly but was accompanied by an incorrect motor response. We had to make certain assumptions in this classification effort, but they were explicit. For example, if in applying a rule to a problem-component 46 the subject spoke "6", we assumed that they incorrectly applied the last rule (if two digits are more than 2 apart, the answer is the last digit) rather than the midpoint rule (if two digits are 2 apart, the answer is the midpoint). We could not categorize 23% of the errors without making what we thought were untenable assumptions.

There were two important findings from this study. First, as we hypothesized, most detected errors were motor errors (subjects said the correct answer but pressed the wrong key). All but 6% of the errors that people detected (pressed the space bar to retake the trial) were motor errors. This indicates that very few cognitive errors were detected, which underscores the importance of studying cognitive slips by skilled performers. Unlike motor errors, there are no observable actions that can be compared to intended actions in the execution of cognitive operations. Consequently, cognitive slips may typically go undetected by the performer, especially when they occur during skilled performance that is fast and lacking in attention control.

The second finding of interest involved errors that occurred in the third operation of 3-rule sequences that were new (transfer) sequences. These errors were especially critical to the strong-but-wrong hypothesis, for new (transfer) sequences differed from old (training) sequences only in this third position. As in previous studies, there were more third position errors on new sequences than old sequences. More importantly, 85%

of the third position cognitive errors in new sequence trials were from triggering the wrong rule rather than misapplied correct rule. This was in contrast to 37% of the errors being wrong-rule errors in the old sequence trials during the transfer session. This suggests that new processing sequences can induce strong-but-wrong errors as hypothesized. It should be noted though that not all errors were of this type. New processing sequences also induced some errors that were misapplication of correct rules.

Implicit Versus Explicit Memory Representation of Sequence Knowledge

Several forms of evidence from previously described experiments were consistent with the interpretation that the sequence knowledge responsible for the errors was procedural and implicit rather than declarative and explicit in nature. First, in the studies that assessed the detection of errors using the retake option, the majority of errors were undetected by the performer. Although no error detection measure was used in the first experiment with simple number reduction, the persistent high rate of errors during 10 transfer blocks suggested that participants in this experiment also were unaware of their mistakes. If sequence knowledge were explicit, one would expect errors of retrieval failure or confusion in how to apply the knowledge. Both of these error types would probably be detectable to some degree. It seems more likely that undetected errors stem from the inappropriate firing of implicit procedural memory representations that are executed with little conscious effort.

Second, for highly skilled learners, error responses to new transfer sequence trials in the previous experiments were fast. These response times were similar to correct response times for old processing sequences, for which we assume the high skill participants relied on procedural memory. If new sequence errors were due to retrieval failures or confusion, response times should have been longer.

Although these results were consistent with the interpretation that sequence knowledge was implicit and procedural in nature, the evidence was indirect. In the current experiment (see Woltz, et. al 2000 for more details), we tested the procedural and implicit nature of processing sequence memory more directly. We practiced individuals in a version of number reduction that resembled the high skill training condition of Experiment 2. However, instead of performing transfer trials, participants judged whether the transfer trials were old or new. In one condition, new transfer trials represented new sequences. In another condition, new transfer trials represented new instances of old sequences. All trials in the final session probed participants as to whether they had experienced the trial in previous training sessions. On half of these trials, participants responded with their number reduction solution prior to the recognition probe, and on the other half they were presented with only the recognition probe. We included performance trials to investigate the relationship between recognition and performance, and we included some trials with recognition-only in the event that performing the number reduction trial first affected subsequent recognition accuracy.

Several comparisons in the data tested whether sequence and instance effects seen in the previous studies could be dependent on declarative or explicit knowledge rather than implicit procedural knowledge. First, we tested whether participants were able to discriminate between old and new sequence trials during transfer. If sequence knowledge

was explicit, then old-new recognition in the sequence condition should have been well above chance. Second, we tested whether sequence recognition was better than instance recognition. If sequence and instance effects in performance measures were both due to explicit memory retrieval, then sequence recognition should be better than instance recognition because sequence effects were larger than instance effects. Third, we evaluated whether old sequence trials that were correctly recognized were performed with shorter latency and fewer errors than unrecognized old sequence trials. This should be the case if performance facilitation due to sequence knowledge depended on explicit recall of the sequence. Finally, we tested whether individuals who showed the greatest old-trial performance facilitation relative to new-trial performance had the highest old-new recognition accuracy. A failure to find these predicted effects would be inconsistent with an explicit memory explanation of sequence and instance learning, and consistent with a conclusion that the knowledge is procedural and implicit in nature.

With regard to the performance latency of recognized old sequences, more than one interpretation is possible if these trials were performed faster than unrecognized old sequences. As noted above, it could indicate that the previously observed old-new sequence differences are attributable to explicit recognition of old sequences. However, it could also indicate that participants use perceptions of performance fluency to make recognition decisions. This phenomenon has been demonstrated in other cognitive tasks (Johnston, Hawley, & Elliott, 1991). If this occurred, we would expect it to be most prevalent in the trials that demanded trial performance prior to recognition. However, we could not prohibit participants from solving the problems prior to recognition judgments in the recognition-only trials. Because they predict the same performance-recognition pattern for old sequence trials, the later interpretation (performance influences recognition) is difficult to distinguish from the former (recognition influences performance).

Table 1 presents the mean recognition data by condition. First note that in both the instance and sequence recognition conditions, there was a considerable difference in accuracy for old and new trials, $p < .05$. This difference reflects a bias toward calling trials old. This bias was stronger in the sequence recognition condition, as reflected by a significant interaction between trial type (old vs. new) and recognition condition, $p < .05$. This is consistent with the hypothesis that participants would often misperceive the new sequences as old, thus leading to strong-but-wrong procedures during performance.

Overall discrimination between old and new trials was represented by the $\underline{d}$' statistic. As can be seen in Table 1, the mean $\underline{d}$' values for both groups were only slightly greater than zero. However, the difference from zero (i.e., chance discrimination) was statistically significant for both the instance and sequence recognition groups, $p < .05$. Thus, participants had better than chance discrimination between new and old trials.

We also tested the difference in discrimination ability between the two recognition groups (sequence vs. instance). The sequence group had slightly better discrimination as indexed by the $\underline{d}$' statistic, but this difference was not statistically significant, $p > .10$. Although sequence knowledge had a reliably greater impact on performance than instance knowledge in the current experiment, Experiment 2 reported earlier, and Woltz et al. (1996), there was only a non-significant trend for explicit recognition of sequences to be better than recognition of instances.

Table 1

Mean Recognition Accuracy by Condition

Variable	Instance Recognition		Sequence Recognition	
	$\underline{M}$	$\underline{SD}$	$\underline{M}$	$\underline{SD}$
Old Trial Accuracy (%)	66.53	17.00	82.28	15.91
New Trial Accuracy (%)	39.86	17.87	23.30	15.59
Old-New Discrimination ($\underline{d}$')	.19	.20	.30	.40

Next, we examined performance differences between old trials that were correctly identified as old and those identified as new. If performance facilitation on old trials depended on recognition, then recognized old trials should have been performed faster and with fewer errors compared to unrecognized old trials. However, as noted earlier, such a pattern would also be consistent with an interpretation that recognition decisions were made on the basis of perceived performance fluency. Figure 8 presents mean error and latency data by trial condition[ii]. Although our primary interest was in the performance of old trials, new trials were also included for reference purposes.

As seen in Figure 8, differences between recognized and unrecognized old trials were small in the instance recognition condition. The difference was not statistically significant in latency or errors, $p > .10$. Thus, there was little or no association between performance and recognition in the instance recognition condition.

The picture was slightly different in the sequence recognition condition. The difference between recognized and unrecognized old trials was relatively small but statistically significant in response latency, $p < .05$. The difference was not statistically significant in errors, $p > .10$. Thus, there was some association between performance and recognition in the sequence recognition condition.

It is difficult to distinguish between the two plausible interpretations of this association (recognition affecting performance versus performance affecting recognition). Both predict the latency difference between recognized and unrecognized old trials. In addition, both predict a similar difference between new trials perceived as old versus new. This prediction is most straightforward for the interpretation that performance fluency affects recognition: New trials that are performed more quickly would be designated as old trials. For the interpretation that recognition affects performance, it is conceivable that new trials incorrectly recognized as old would be performed faster because a response for a similar old trial would be executed quickly. However, this
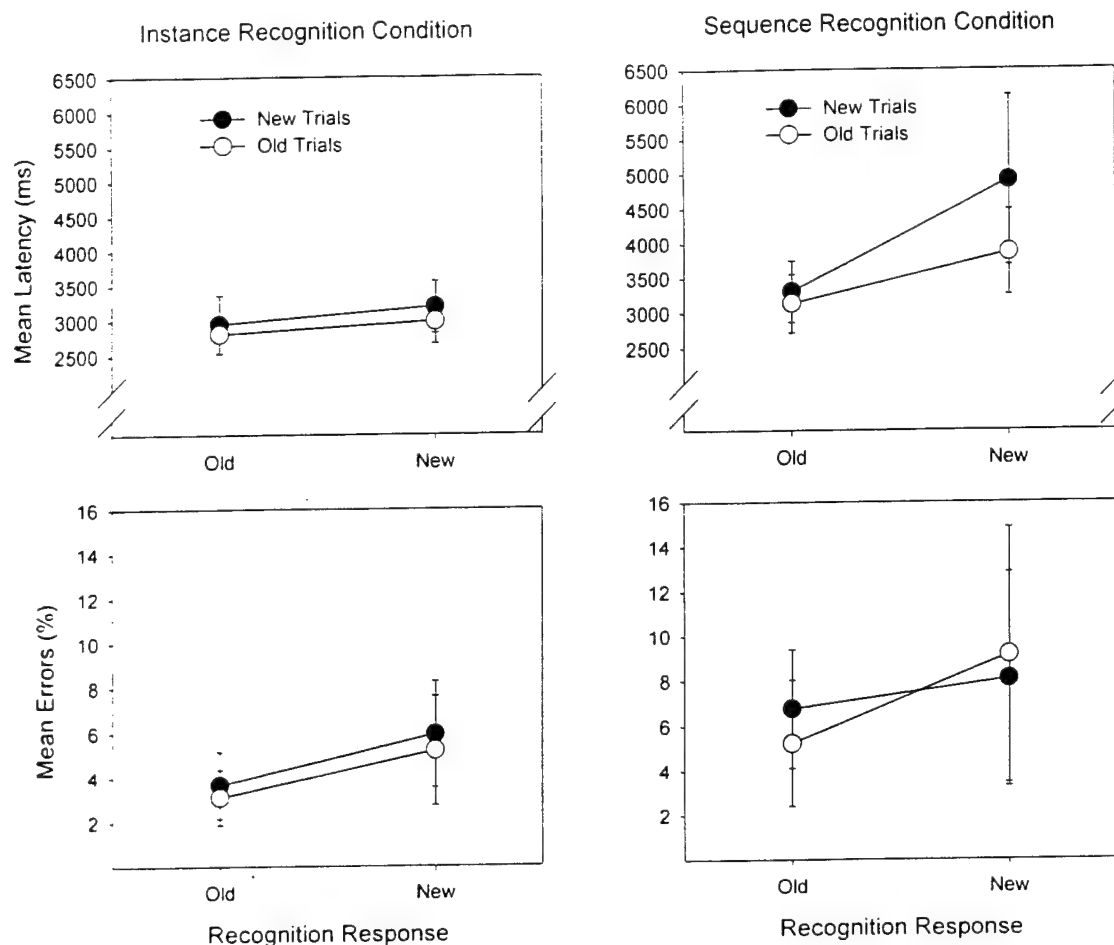
Figure 8: Mean Latency and Errors by Recognition Response

prediction must be coupled with the prediction of increased errors on new trials recognized as old. As seen in Figure 8, this latency difference predicted by both interpretations was found for new trials, p < .05. The error difference between new trials recognized as old versus those recognized as new was not significant, p > .10. In fact, the non-significant trend went in the direction opposite of that predicted by the recognition-affects-performance interpretation. There were slightly fewer errors on new trials when they were incorrectly recognized as old. So, as a whole the data in Figure 8 are more consistent with the interpretation that the association between performance and recognition is due to participants using performance fluency perceptions to assist in old-new recognition judgments.

Finally, we tested the association of performance and recognition by correlating these measures over individuals. If performance facilitation was dependent on explicit memory processes, then there should have been a significant correlation between the accuracy of recognition ($\underline{d}'$) and the magnitude of performance facilitation in old relative to new trials. We indexed performance facilitation as residuals from regressing old trial

latency on new trial latency. Positive residuals reflect individuals who took longer on old trials than predicted from their new trial latency. Negative residuals reflect individuals who were faster than predicted by their new trials performance[iii]. For the instance recognition condition, the correlations was $r = -.15, p > .10.$ For the sequence condition, the correlation was $r = -.12, p > .10.$ Both correlations were sufficiently low to be inconsistent with the hypothesis that performance facilitation on old trials was linked to recognition of those trials[iv].

Taken as a whole, the results of this experiment suggest that knowledge for processing sequence, which impacts both performance latency and errors, represents an implicit rather than explicit memory process. First, although recognition performance was better than chance in both instance and sequence recognition conditions, it was relatively poor. The mean percentage correct on old-new decisions was just 53% for both groups (50% being chance). Second, sequence effects were greater than instance effects in the performance data (also see Experiment 2 reported here and Woltz et al., 1996), but sequence recognition was not reliably better than instance recognition. Third, although participants in the sequence group were faster in performing old trials that were correctly recognized as old, this did not appear to reflect an influence of recognition on performance. Instead, the combination of error and latency data were more consistent with the interpretation that perceived performance fluency influenced the old-new recognition decisions. Finally, there was a low and non-significant correlation between the magnitude of sequence effects and recognition accuracy.

The previous experiments demonstrated that processing sequence knowledge was responsible for near transfer errors under certain task conditions. The purpose of the current experiment was to determine whether the influence of the sequence knowledge was likely to be explicit in nature (e.g., using a conscious strategy based on sequence recognition), or implicit (e.g., the triggering of a strong procedural memory without awareness). This issue is difficult to resolve completely, both because of methodological difficulties and because both forms of knowledge may operate in tandem to some extent. However, the bulk of evidence reported here favors the implicit memory interpretation.

Contrasting Composition, Rule Transition, and Associative Chain Representation

In two experiments, we contrasted three alternative memory representations for processing order information. Composition or chunking constitutes a representation of each complete sequence of operations in a skill as a whole unit (see Anderson, 1983, 1987). Composed sequences are thought to execute as a whole in an all-or-nothing fashion. The existence of composition in skill representation has been controversial (Anderson, 1993; Carlson & Schneider, 1989; Carlson, et al., 1989), although it remains a topic of consideration by some (Charness & Campbell, 1988; Frensch, 1991, 1994). A second form of sequence representation is simple memory for rule transitions. This possibility allows only for dyad transition knowledge, rather than complete sequence representation. Finally, sequence memory could be represented as associative chains that provide anticipatory priming for subsequent operations in a chain of any length. This differs from composition in that sequence knowledge does not execute in an all-or-nothing fashion for the entire sequence. Instead, each step is primed in sequence by prior

steps. This differs from simple rule transition representation in that familiar steps at the end of a sequence show performance facilitation only if the beginning of the sequence was familiar also. That is, priming for later steps depends on the entire chain of steps being familiar. Simple transition memory produces performance facilitation for any familiar step, regardless of the familiarity of the entire chain of steps.

These three representations make similar prediction in most cases, but as noted there are specific conditions under which they differ. We conducted two experiments in which the different representation theories made partially contrasting predictions.

Experiment 1. In Experiment 1 of this series we taught 67 participants a version of the number reduction skill in which each problem required the sequential application of three rules. The three-rule sequences were drawn from a population of all possible orderings of four computational rules. During training, participants saw only a subset of the possible orderings of the rules. During transfer, participants saw all the possible orderings. Presentation of rules during training was counterbalanced such that each rule was presented an equal number of times in each serial position. This equated the strengthening of individual rules by serial position.

The question of interest was addressed during transfer. All transfer items were different from training items in terms of their surface structure. Thus instance effects were equated for old and new sequences in this experiment. Questions of sequence representation were addressed by varying the similarity of rule combinations in transfer to the original training trials. Transfer rule sequences could match training rule sequences in either the first two rules (e.g., A-B of the rule sequence A-B-C; we refer to this as a *first rule dyad match*), the second two rules (e.g., B-C of the rule sequence A-B-C; we refer to this as a *second rule dyad match*), both the first two rules and the second two rules, but not the all three rules (the first and second rule dyads match, but not the rule triad; this was possible because a transfer item could match the first rule dyad of one training sequence and the second rule dyad of a different training sequence), or neither the first two rules nor the second two rules (no dyads match). It was also possible to match the rule triad, which implied a match of the first and second rule dyads (these were training rule sequences seen during transfer with new item content). First we consider the predictions of each theory of sequence representation. We made predictions about new transfer trial performance relative to performance on training trials (i.e., triad matches). These predictions are summarized in Figure 9.

A dyad transition model of sequence representation makes the simple prediction that latency and errors will increase for new transfer trials to the extent that dyad transitions differ from those in training sequences. As shown in the left two panels of Figure 9, when both dyads are new, latency and errors will be greatest. When only one dyad is new, latency and errors will be increased to the same extent, regardless of which dyad is new. Of particular importance, latency and errors for trials with two old dyads (but a new triad) should not differ from latency and errors for training trials.

A composition model of sequence representation makes different predictions for both latency and error compared to the dyad transition model. As shown in the upper middle panel of Figure 9, it predicts longer latency when the first dyad differs from training sequences. A new first dyad prevents the composed representation from firing, so it is irrelevant whether the second dyad is old or new. In contrast, an old first dyad is assumed to be sufficient to trigger the composed production, so latency for an old 1st
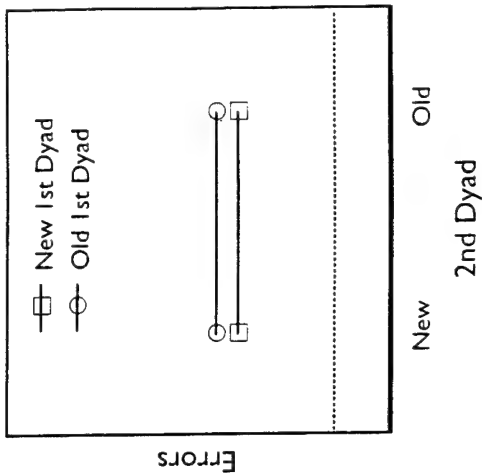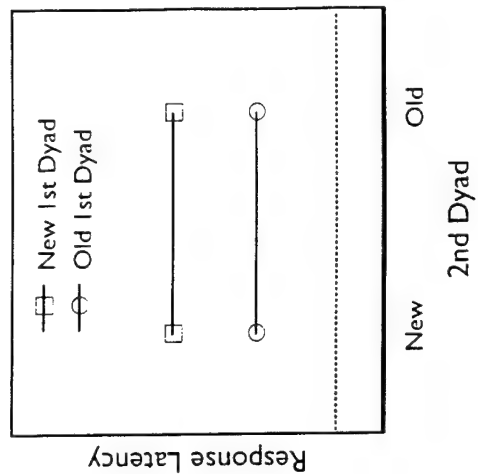
Dyad Transition
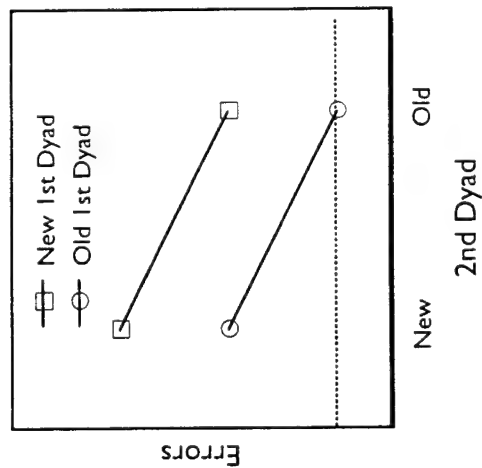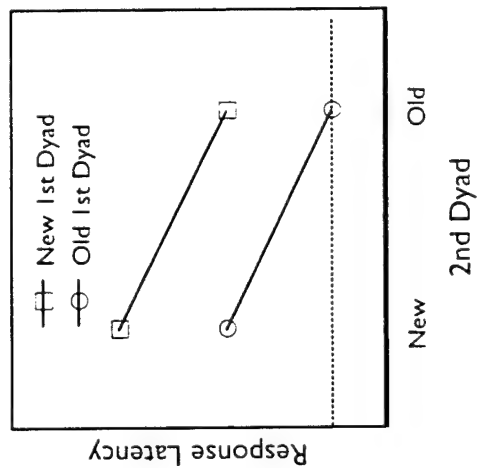
Composition
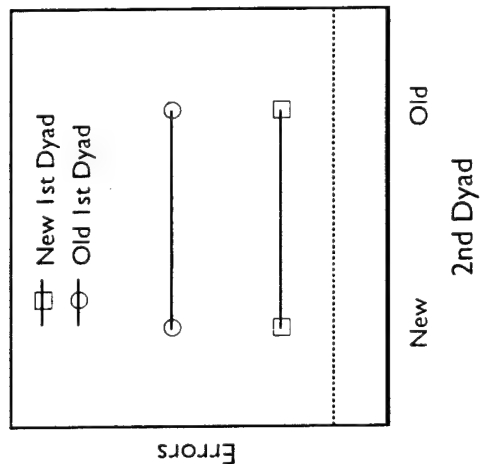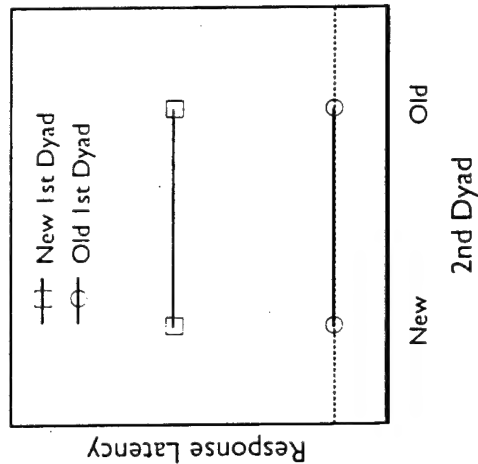
Associative Chain

Note: Dashed lines represent performance on training sequences (i.e., triad matches) presented during the transfer phase. Differences in model predictions for new transfer trials reflect the expected relative difference between trial conditions, not the absolute magnitude of these differences.

dyad trial should not differ from that of training sequences that rely on the same representation. With respect to errors, a composition model makes the unique prediction that there will be more errors in old 1st dyad trials than in new 1st dyad trials (see the lower middle panel of Figure 9). As noted above, an old first dyad is expected to invoke the all-or-nothing execution of the complete sequence representation. This should produce a high rate of 'garden path' errors. Note that new 1st dyad trials are also expected to produce more errors than training sequences, but not as many errors as old first dyad sequences. The errors associated with new 1st dyads reflect the lower reliability of reverting to 'weaker' representations (e.g., declarative or initial procedural knowledge for individual rule components).

Finally, an associative chain model of sequence representation makes predictions about transfer performance that are distinct from either dyad transition or composition models (see the right two panels of Figure 9). As with composition, any trial that begins with a new dyad is expected to have the longest latency. In contrast to composition, an associative chain model predicts that new sequences in transfer that begin with old 1st dyads will produce longer response latency than will training sequences. On these trials, there is partial facilitation from the initial match with the associative chain representation, but latency is subsequently increased relative to training sequences by the need to revert to other representations to complete the trial (i.e., declarative knowledge or procedural representations of individual component rules). Furthermore, because associative chains are not all-or-nothing in their execution, there is no prediction of high error rates for trials that begin with an old 1st dyad but end in a new way, as was the case with composition. We would expect some 'garden path' errors on trials that begin with a familiar dyad but end differently. However, there is no reason to expect the frequency of these errors to be higher than errors due to reverting to weaker representations.

It is possible to conceive of more complex hierarchical models in which multiple levels of sequence representation co-exist. For example, both dyad and triad representations could exist, and dyad representations could be independent of serial position within the rule sequence. These models were not tested in this study. We limited our predictions to simple versions of these models where the dyad and its serial position within the sequence are represented jointly.

Figure 10 presents the results of this experiment. The upper left panel (a) shows the latency and error data during the four training sessions. The training data reflect typical skill acquisition learning curves. The upper left panel (b) shows the mean latency by trial type during transfer. This panel should be compared to the predicted outcomes in the top three panels in Figure 9. Visual inspection suggests that the data conform to the predictions of the associative chain model of representation. Statistical analyses confirmed this. Old 1st Dyad trials differed from both training sequence trials and New 1st Dyad trials ($p < .05$). However, New and Old 2nd Dyad trials did not differ beyond what would be expected by chance ($p > .10$).

The error data from this experiment also conformed to the predictions of the associative chain model. The lower left panel (c) of Figure 10 shows the detected errors by trial type during transfer. This merely shows that few errors were detected, and that the number of detected errors did not differ between training and transfer sequences. The lower right panel (d) of Figure 10 shows the undetected errors by trial type. This panel should be compared to the three lower panels of Figure 9. Again, statistical analysis
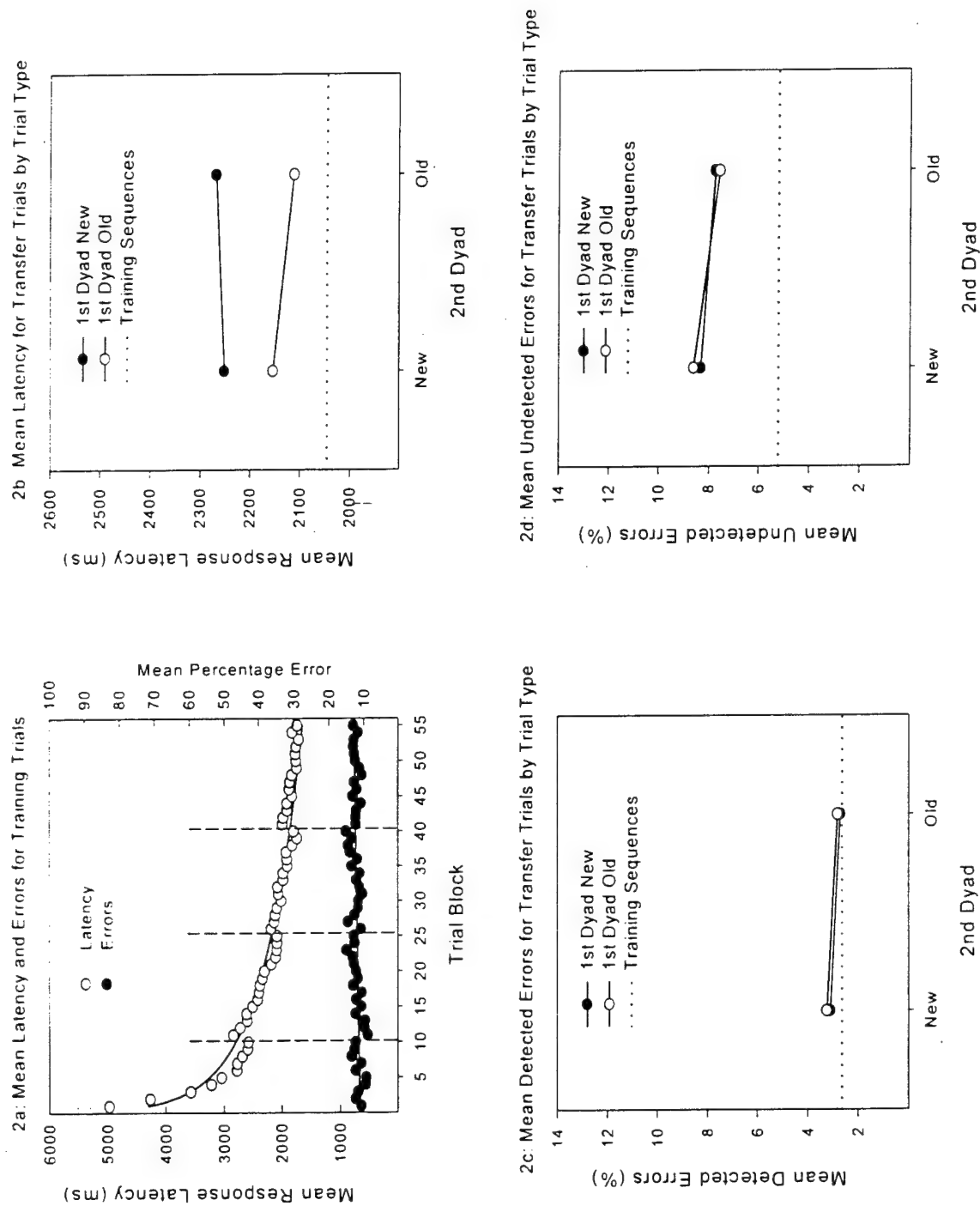
Figure 10: Results of Experiment 1 in Series Contrasting Forms of Sequence Representation

confirmed the similarity to the predictions of the associative chain model. There were more trials on new transfer sequence trials as a whole compared to training sequence trials (p < .05). Furthermore, there was no difference in the number of undetected errors between either old and new 1st dyad trials or between old and new 2nd dyad trials (p > .10).

Experiment 2. The pattern of latency and error data in the various transfer conditions of Experiment 1 of this series led us to reject a composition and dyad transition models of sequence representation in favor of a complex associative chain model. However, the design of the experimental task may have unduly disadvantaged the composition model. Composition might be more likely to occur in skills that require fewer sequences to be learned. Also, composed sequences might be triggered during transfer in the manner predicted for Experiment 1 only when there is a close match between training sequence surface structure and transfer sequence surface structure. In Experiment 1, digit strings presented during transfer were always different from those presented during training, even when the sequence of operations was identical to those from training.

The evidence thus far suggests that sequence memory has a degree of generality. That is, memory for the sequence of processing operations facilitates performance even with new surface structure of individual trials (i.e., new data on which the sequence of operations executes). In Experiment 1, we contrasted composition and other models under conditions that assumed data-general sequence representation. However, it is not clear that the composition mechanism is capable of handling such generality. Carlson and Schneider (1989; Carlson, Sullivan, & Schneider, 1989) argued that for a composition mechanism to work, it logically must incorporate data-specific aspects of the particular instance viewed. In tasks such as number reduction, the output of one step determines the input to the subsequent step. Furthermore, intermediate step solutions determine which subsequent operations are applicable. Real-time processing adaptations that depend on intermediate solutions are inconsistent with the notion of all-or-nothing execution of a composed set of steps. Under this view, composition should not be possible, unless instances were consistent in both training and transfer.

Anderson (1989) disagreed with the need to retain item surface structure within composed productions. He allowed variables to be composed in place of specific intermediate results, thus allowing for instance-independent sequence memory. While the finding of instance-general sequence memory effects would seem to support Anderson's position, the data from Experiment 1 were otherwise inconsistent with a composition explanation. It should also be noted that Anderson (1993) dropped the composition mechanism in a later version of the ACT theory.

Experiment 2 was designed primarily to assess whether transfer performance data conforms to general predictions of the composition model when new sequence transfer trials resembled training trials in the first dyad and in the first three digits. The composition model predicts that when new sequences begin like training sequences in the first dyad, and when they are identical to a training instance that had been repeatedly practiced, latency will be as fast as that for old training instances and undetected errors will be substantially higher than any other trial condition. In addition, if all-or-nothing execution of composed productions is triggered by this nearly complete match of training

stimulus conditions, the latency of undetected errors should not differ from the latency on training sequences performed correctly.

A total of 51 participants performed a version of number reduction similar to previous experiments. During training participants practiced four rule sequences, with each sequence being represented by 12 instances per sequence. During transfer, participants received a total of eight rule sequences, with each rule sequence being represented by 12 instances per sequence.

Of the eight transfer sequences, four were old (i.e., seen during training) and four were new. The four old sequences were represented by two categories of instances: old instances seen during training (designated old/old), and new instances (designated old/new). The four new sequences matched the old sequences in the first rule dyad (A-B), and were also represented by two categories of instances: instances that matched old instances in the first three digits (e.g., 4656, which matches the old instance 4659 in the first three digits [though these represent different sequences]; these were designated new/old), and instances that did not match old instances (designated new/new).

Which four sequences were used as training and which four were used as new sequences during transfer was counterbalanced over participants. This allowed us to measure whether our effects were strongly determined by the particular rule sequences and instances used.

Figure 11 shows the mean latency and error data from this experiment. The upper left panel (a) presents the mean training data. As with previous experiments using the number reduction task, the data conform to typical skill acquisition learning curves. The upper right panel (b) contains the mean latency data by trial type for transfer trials. Two trial type contrasts were of interest. First, the contrast of old sequence/new instance versus new sequence/new instance tested the presence of data-general sequence memory. This contrast was statistically significant ($p < .05$) with old sequences being approximately 170 ms faster than new sequences. As in Experiment 1 of this series, there was strong support for facilitation due to the same operations being applied in the same order, even though the data being operated on was new. Second, the contrast of old sequence/old instance versus old sequence/new instance tested the role of instance memory beyond that of sequence memory -- that is, facilitation due to identical item content or surface structure in training sequence trials. This contrast was also statistically significant ($p < .05$), with old instances being approximately 105 ms faster than new instances. Clearly, some portion of participants' performance on training sequences was instance based. This was consistent with previous research using the current task paradigm (Woltz et al., 1996; Woltz et al., 2000), as well research using other tasks (Carlson & Lundy, 1992; Logan, 1988).

The lower left panel of Figure 11 presents median undetected errors in transfer as a function of trial type. Undetected error rates ranged between 3.0% and 8.5% across trial types. The measure of sequence memory, the contrast of old sequence/new instance versus new sequence/new instance, was statistically significant ($p < .05$). So, there was evidence for processing sequence facilitation in the undetected error data, as there was in the latency data. The measure of instance based facilitation, the contrast of old sequence/old instance versus old sequence/new instance, was also statistically significant ($p < .05$). Again, as in the latency data, performance was to some degree instance based.
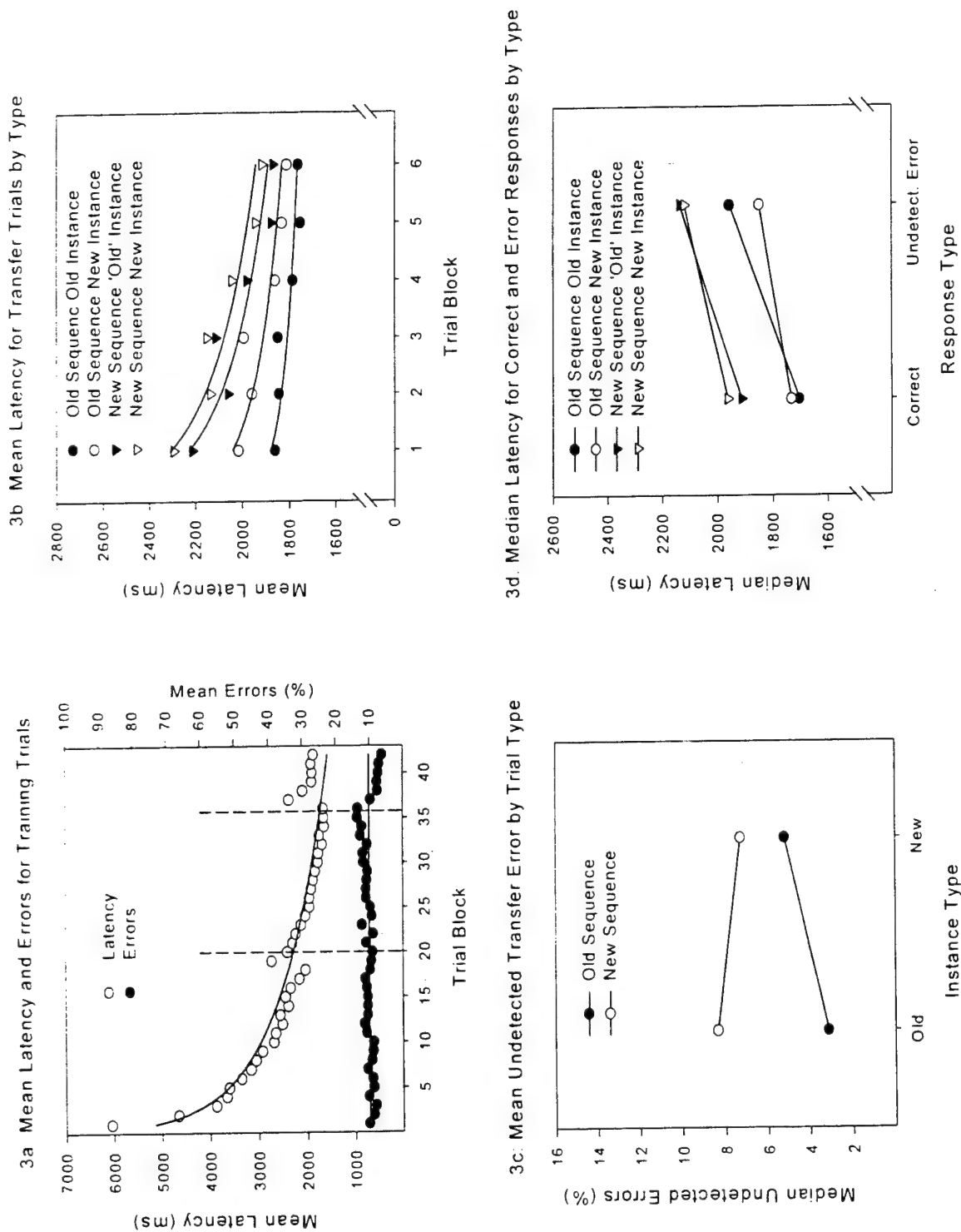
3b Mean Latency for Transfer Trials by Type

● Old Sequenc Old Instance
○ Old Sequence New Instance
▶ New Sequenc 'Old' Instance
▽ New Sequence New Instance

Mean Latency (ms)

2800
2600
2400
2200
2000
1800
1600
0

Trial Block

1 2 3 4 5 6

3d Median Latency for Correct and Error Responses by Type

● Old Sequence Old Instance
○ Old Sequence New Instance
▶ New Sequence 'Old' Instance
▽ New Sequence New Instance

Median Latency (ms)

2600
2400
2200
2000
1800
1600

Correct   Undetect. Error

Response Type

3a Mean Latency and Errors for Training Trials

Mean Errors (%)

100
90
80
70
60
50
40
30
20
10

○ Latency
● Errors

Mean Latency (ms)

7000
6000
5000
4000
3000
2000
1000

Trial Block

5 10 15 20 25 30 35 40

3c Mean Undetected Transfer Error by Trial Type

● Old Sequence
○ New Sequence

Median Undetected Errors (%)

16
14
12
10
8
6
4
2

Old   New

Instance Type

Figure 11: Results from Experiment 2 in Series Contrasting Forms of Sequence Representation

Composition made the unique prediction that a "partial match" of the instance stem (the first three digits) in new sequences would cause the firing of an incorrect an "old" composed production developed during training. This would result in a higher undetected error rate in the new sequence/old instance condition versus the new sequence/new instance condition. Furthermore, these undetected errors should have latencies that are equivalent to correct responses in the old sequence/old instance condition. The new sequence/old instance versus new sequence/new instance contrast for undetected errors was not statistically significant, ($p > .10$). In addition, we considered the absolute number of errors made in the new sequence/old instance condition. The error rate here was 8.33%, which was at best moderate. If this condition represented the firing of composed productions due to a partial match of the enabling conditions, we would have expected a much higher error rate. This data seems more consistent with an associative chain representation of sequence information.

The lower right panel of Figure 11 presents the latency data for undetected errors and correct responses in Experiment 2 as a function of item category. The number of observations per condition is 28, rather than 51, because some subjects made no undetected errors in some conditions. As can be seen in this panel, there was a considerable difference in latency between undetected errors in the new sequence/old instance condition (Mdn = 2136 ms) and correct responses in the old sequence/old instance condition (Mdn = 1704 ms). A test of the contrast was statistically significant ($p < .05$). Thus, analysis of both the undetected error data and the latency data failed to support the predictions of the composition model.

The results of this experiment, while inconsistent with a composition representation of sequence information, were consistent with an associative chain representation of sequence memory. There was clear support for both sequence-based and instance-based memory effects in the latency and undetected error data. Both composition and associate chain representations predict such effects. However, additional predictions made by the composition model (i.e., all-or-none firing of composed productions, triggered by a partial match of the production's enabling conditions) were not supported in either the undetected error data or the latency data for these errors. Thus, as in Experiment 1 of this series, the data are more consistent with an associative chain representation.

Generalization of Sequence Memory and Associated Errors Across Skills

It was deemed important to investigate the impact of sequence memory on near transfer performance using additional skill tasks to ensure that the phenomena observed in number reduction experiments were not task-specific. In previous research in our lab, we found evidence of sequence representation in the number reduction task (Thurstone & Thurstone, 1941; Woltz, Bell, Kyllonen, & Gardner, 1996) and a more real-world like computational skill (see Elio, 1986). However, we did not find evidence that sequence memory was instrumental in another skill task (see Woltz, 1988). It now appears from our data as well as others (Lundy, Wegner, Schmidt, & Carlson, 1994) that step dependence is important for the representation of sequence knowledge. That is, steps must depend on output of previous steps (cascading). Without this, multiple steps can

ultimately be processed in parallel rather than sequentially. This explanation is difficult to test in a controlled fashion, but at a minimum it is important to verify that sequence representation is important to additional skill tasks with step dependence. It is also worth noting that complex real-world skills often have sequence dependence (e.g. computational and diagnostic skills).

All of our initial evidence regarding theoretical issues of skilled performance errors came from experiments conducted with the number reduction task. This experimental task had advantages for addressing many questions, including the rapid acquisition of the skill by subjects and the tasks ability to handle meaningful manipulations without changing the general nature of the skill. However, our concern was that the empirical phenomena and the theoretical interpretations of them that resulted from number reduction experiments would not generalize to other skill tasks, especially those that more closely resemble real-world skills. Although generalization is often the last issue to be addressed in experimental research, we felt that it was important to investigate the generality of errors due to sequence memory in several sequential skill tasks.

Math Problem Solving

Math problems represent a common domain of cognitive skills that often involves specific sequences of operations. We developed a series of 3-step math problems using the addition, subtraction, multiplication, and division operators. In the following example, $((7 \times 9) + 1) / 8$, the multiply, add, and divide operations must be applied in that order. The problem $((3 \times 5) + 3) / 2$ is a different instance of the same processing sequence.

The demonstration of sequence memory in this task was important for two reasons. First, subjects have extensive experience with component operations in this task (i.e., the addition, subtraction, multiplication, and division facts). It is possible that practicing sequential applications of over-learned operations may not lead to sequence memory that has much effect on performance. Second, individuals have presumably had a history of applying these operations in various sequences. Thus, while we have found sequence memory effects in novel skills for which people have no prior experience, it is another question whether this effect can be shown for more common skills that have prior histories of varied sequential applications. This study then has importance to whether sequence memory findings generalize to many real world skills for which acquisition of all components does not take place in a highly controlled laboratory setting.

We manipulated the acquisition of sequence memory by restricting the number of operator sequences seen during a training phase. Then we introduced additional sequences of the same operators during transfer, comparing the old and new sequences to estimate the impact of sequence memory on both latency and errors.

In this study each participant (n=31) practiced problems representing four 3-operation sequences. Frequency of operation by position was balanced within each set of four sequences, so sequence learning effects could not be attributed simply to serial position information. There were 16 unique instances per sequence seen during training. Each participant performed three sessions of training (768 total trials), and one session of transfer. During transfer, half of the trials were training sequences. Of these, half represented old instances (Old-Old) and have represented new instances not seen in

training (Old-New). The other half of trials represented new training sequences that necessarily were new instances (New-New).

During training sessions, performance times improved gradually according to a power function. Mean initial performance per trial approximated 6 s and was reduced to approximately 3 s at the conclusion of the training sessions. Performance improvement could be due to any combination of (a) general task familiarity, (b) arithmetic component tuning, (c) problem instance representation, (d) and processing sequence representation.

Figure 12 shows transfer mean latency data by trial block (24 trials per block). The difference between Old-Old and Old-New trial performance represents the degree to which skill performance was instance-based. As can be seen in the figure, this was a substantial effect. This finding corresponded to subjects anecdotal reports that they were rusty on retrieving many math facts and became more fluent with practice. The difference between New-New and Old-New trial performance represents the degree to which skill performance was sequence-specific. Both of these trial types were new instances, so any difference can be entirely attributed to familiarity of the sequence of operations. As can be seen in Figure 1, this effect was smaller than that for instance representation but still noteworthy (and statistically significant, $p < .05$). There were more errors in transfer on both new instance and new sequence trials compared to the old-old training trials ($p < .05$). There was also a nonsignificant trend of more errors on new sequence trials compared to new instance trials. Additional data are being collected in an attempt to further understand the error effects.

Given the familiar nature of all components of this skill, these findings underscore the generality of processing sequence representation. Although the magnitude of its effect relative to that of processing instance representation was smaller than in previous studies using unfamiliar skills, processing sequence memory was still instrumental.

Logic Gates

Logic gates refer a set of boolean operators that are used in circuit design and evaluation The acquisition of skill in evaluating logic gates has been studies by previous researchers (e.g., Kyllonen & Stephens, 1990), and experimental versions of the skill appear to represent another class of real-world sequential cognitive skills that can be studied in the laboratory. This skill differs from previous skill tasks investigated in our laboratory in that it requires the application of binary transformation rules represented by unfamiliar graphic symbols.

In our studies each participant practiced problems representing 3-gate sequences. Figure 13 shows an example problem in which a subject would enter a response (0 or 1) representing the output signal for each gate in the sequence, moving left to right. The first gate (AND) would produce an output of 0. The second gate (OR) would combine the output from the AND gate with the new input and produce an output of 1. The third gate (NAND) would combine the output from the OR gate with the new input and produce an output of 1. Thus, the correct responses to this problem would be 0, 1, and 1.

Similar to our research with other sequential skills, we manipulated the acquisition of sequence memory by restricting the number of gate sequences seen during a training phase. Then we introduced additional gate sequences during transfer, comparing the old and new sequences to estimate the impact of sequence memory on both latency and errors.
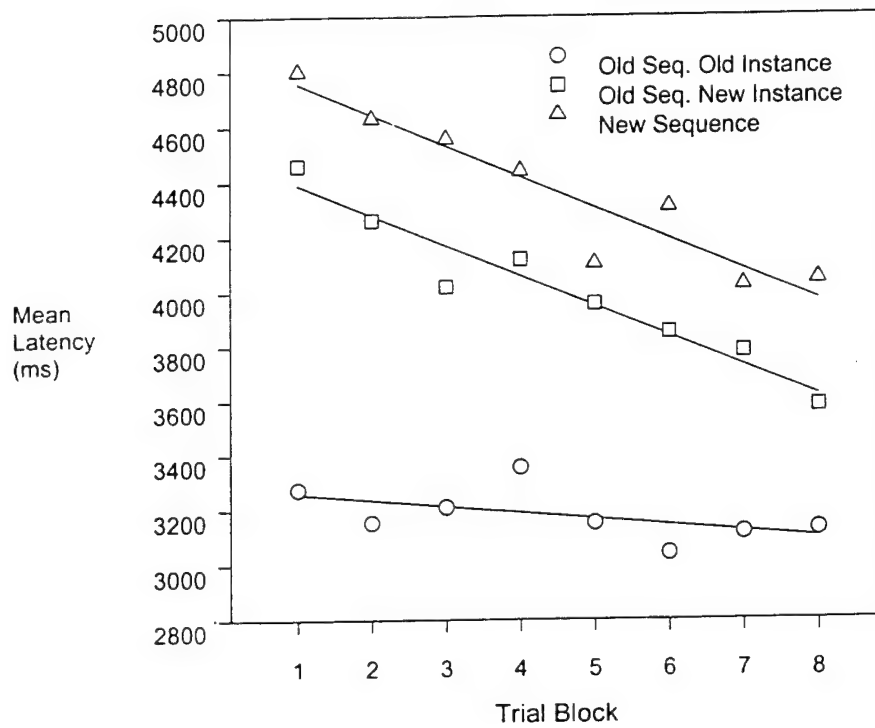
Figure 12: Transfer Latency by Trial Type and Block for Math Experiment
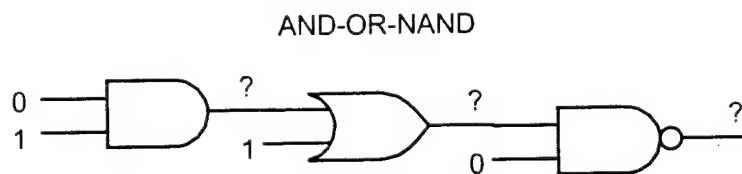
AND-OR-NAND

Figure 13: Example Logic Gates Problem

We conducted several experiments investigating the logic gates skill. We summarize two of them here. These experiments are reported in more detail elsewhere (Woltz, Gardner, & Gyll, in preparation).

Experiment 1. In this experiment, we used a simple version of the logic gates skill. Subjects were trained with a set of only two distinct sequences that balanced the frequency of each of two gates were used at each of the three positions (e.g., *And-Or-And* and *Or-And-Or* were used for some subjects and *And-Or-Or* and *Or-And-And* were used for other subjects). Each 3-step sequence was represented by 6 different instances, where instances differed in the binary input at each step. Two experimental groups practiced the 12 instances during a training phase. The groups differed only with respect to the amount of practice that they had prior to a transfer phase. Both groups had the same amount of instruction regarding the component gates, but one group had fives times as much practice in evaluating the gates in sequence.

The trials in the transfer phase included all of the previously practiced instances of the two training sequences (old sequence trials), and instances of two new sequences that used the same gates with equal frequency at each position. New transfer sequences were created such that each one matched an old training sequence in the first two gates, but ended with a different gate. Negative transfer in the form of strong-but-wrong errors was expected in the form of high skill participants making more errors than the lower skill participants on new sequence trials. Thus, as had been found in other sequential skills, we expected that the partial match of stimulus conditions to strong memory for processing sequences would result in executing incorrect operations in the final problem step.

We also predicted that if the high skill participants made more errors on new sequence transfer trials, the response latency for these errors would be fast, similar to the latency for correct responses to old sequence trials. Given that we could produce errors in the new sequence transfer trials, this prediction was important to discriminate between weak-method explanations (Anderson, 1989) and skilled memory explanations for these errors.

Figure 13 presents mean data for training trials (blocks 1-20) and transfer trials (blocks 21-30) for both high skill (n = 141) and low skill (n = 121) conditions. Note that the low skill subjects received only Blocks 1, 9, 19, and 20 during the training phase, while the high skill subjects received all 20 blocks. The low skill subjects performed comparable computerized tasks as fillers during the period that high skill subjects completed the additional blocks of logic gates.

As shown in Figure 13, the high and low skill groups had equivalent error rates at the end of the training phase. However, the high skill group had achieved a substantially lower average response time. This pattern was evident in responses to all three problem-steps. During the transfer phase, both groups showed an initial increase in errors for new sequences relative to old ones (p < .05). This new-old difference was greater for the high skill subjects, and the difference persisted throughout the transfer phase for high skill but not low skill subjects. Thus, the findings were consistent with those from other sequential skill tasks. New transfer sequences that initially resembled old ones produced errors. Participants with more practice on the old sequences had more errors on the new ones, suggesting that stronger sequence memory resulted in more negative transfer.
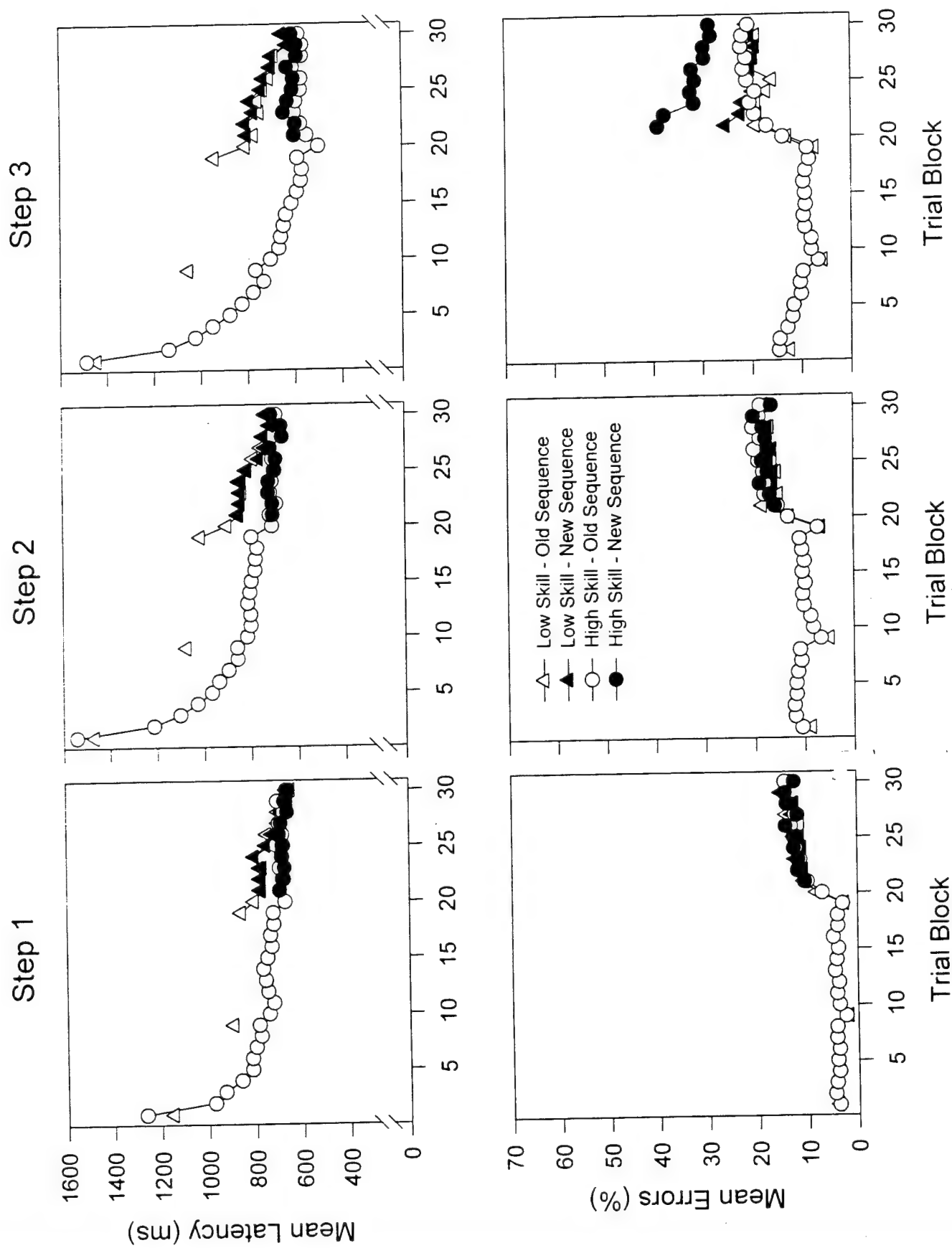
Figure 13: Mean Latency and Errors for Simple Logic Gates Experiment

Experiment 2. In this experiment, we investigated questions of sequence memory and strong-but-wrong errors using a version of the logic gates task that had greater complexity and ecological validity. Each subject practiced problems that represented four distinct 3-gate sequences, using the traditional gate symbols shown in Figure 12. As in previous experiments, frequency of gate type by position was balanced within each set of four sequences, so sequence-learning effects could not be attributed to serial position information. There were eight unique instances for each sequence seen during training (i.e., eight different configurations of input data per sequence).

Each participant performed three sessions of training, and one session of transfer. During transfer, half of the trials were training sequences. Of these, half represented old instances (Old Sequence – Old Instance) and half represented new instances not seen in training (Old Sequence – New Instance). The other half of trials represented new training sequences that necessarily were new instances. Half of these new sequence trials represented a 'new' gate in Step 2 and half contained a 'new' gate in Step 3. The contrast of these two new sequence conditions tested issues concerning the conditions necessary to induce the strong-but-wrong sequence application errors.

Figure 14 shows the mean latency and error data for training trials of this experiment. As with simple number reduction experiments described earlier, there was considerable learning demonstrated in the latency data. In contrast to earlier number reduction experiments, there was also notable learning demonstrated in the error data. The relatively high error rate in early training blocks reflects the increased complexity of this skill task. However, by the end of training, error rates approximated 10%, which was the performance goal in virtually all of our experiments.

Figure 15 shows mean latency and error data by gate order and trial condition for the transfer session. The mean latency data for the three steps closely resembled the latency data by step at the end of training. However, old sequence trials had shorter latency than new sequence trials, $p < .05$. This reflected a general role of memory for gate sequences in the fluency of trial performance. Within old sequence trials, old instance trials had shorter latency than new instance trials, $p < .05$. This difference reflected a role of memory for input configurations that were repeated. There was also a significant interaction between step (2 vs. 3) and type of new sequence (new 2nd Step vs. new 3rd Step), $p < .05$. This reflected the fact that subjects slowed down at the step in which the unfamiliar gate order was introduced.

In the error data shown in Figure 15, there was a general trend for greater errors with each gate simply because errors were cumulative (i.e., an error made in Gate 1 of a problem would generally produce errors in Gates 2 and 3). Three findings were important. First, there were more errors on new versus old sequence trials, $p < .05$. This demonstrates the role of sequence memory in errors within the more complex logic gates task. Second, among old sequence trials, there were more errors in new versus old instances $p < .05$. This demonstrates a smaller but still measurable role of instance memory in skill performance errors. Finally, among new sequence trials, there was a disordinal interaction between step (2 vs. 3) and new sequence type (new 2nd step vs. new 3rd step), $p < .05$. As expected, there were more errors in Step 2 when the sequence deviated from old sequences at this step, and there were more errors in Step 3 when the sequence deviated from old sequences at this step. This finding was consistent with the
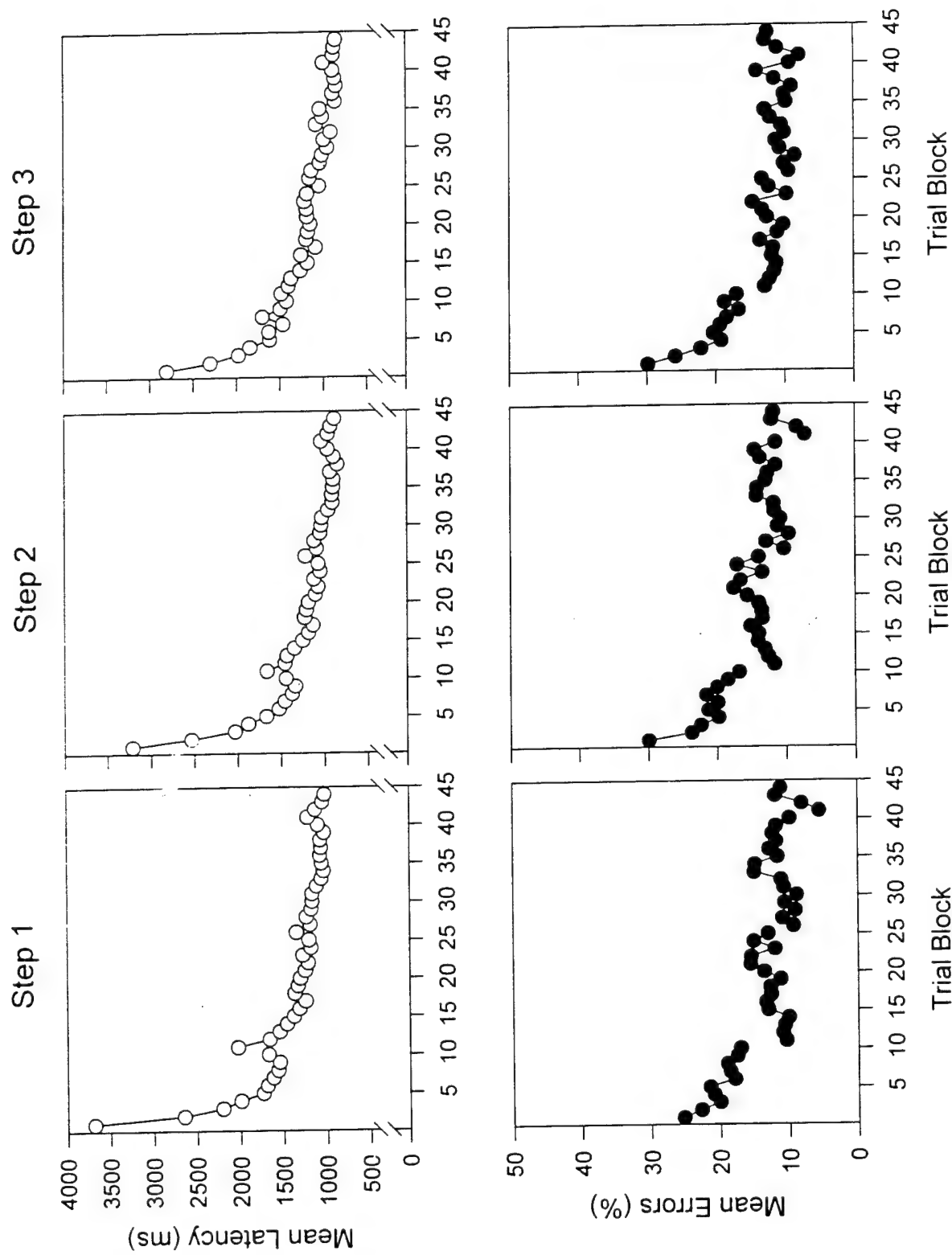
Figure 14: Mean Latency and Errors for Training Trials in the Complex Logic Gates Experiment

notion that sequence memory underlies strong-but-wrong errors, and that the influence of sequence representation on negative transfer errors can occur both early and late in problem sequences of this skill.
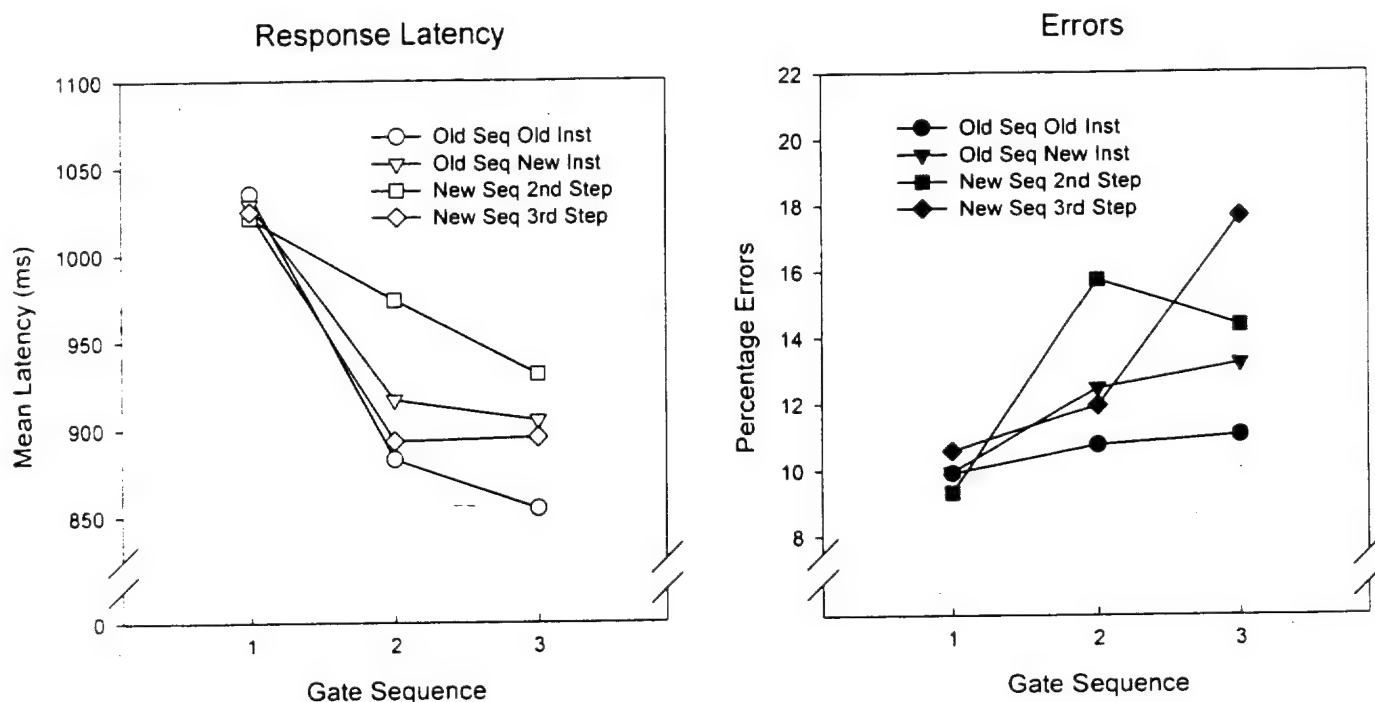
Response Latency   Errors

Figure 15: Mean Latency and Error Data for Transfer Trials by Trial Type and Gate Step

Cognitive Characteristics of Error-Prone Individuals

The understanding of individual differences in error making has two potential payoffs. First, knowledge of cognitive characteristics that are associated with skilled performance errors provides an alternative method of testing theories of cognitive mechanisms underlying the errors (see Underwood, 1975). Specifically, we were interested in investigating through patterns of correlation the potential roles of two cognitive mechanisms: (a) limited working memory capacity and (b) attention disengagement from expected events. Second, there may be practical benefits to personnel classification and personnel training from developing a greater understanding of meaningful learner characteristics.

One study investigated errors in both the number reduction and logic gates skills, and the relationship of these errors to a variety of cognitive measures (see Woltz, Gardner, & Gyll, 2000 for more details). Of primary interest were measures of working memory and attention disengagement. We found that transfer errors were correlated in the two skill tasks, suggesting that error making in skill tasks is a general rather than task-specific

characteristic. In addition, we found that working memory and attention disengagement had different relationships with skilled performance. As shown in other published research (e.g., Woltz, 1988), working memory capacity was associated with early skill performance where learners must acquire and interpret declarative rules that govern performance. Performance after extensive practice no longer demands much working memory capacity, but errors can occur under near-transfer conditions. Our current evidence showed that effective attention disengagement ability was highly related to the ability to perform accurately at later stages of skill acquisition, and especially under transfer conditions where the learner must make minor adjustments in processing according to differing task demands.

One hundred and thirty five participants performed two skill learning tasks and 10 cognitive ability measures over five sessions. The number reduction skill was practiced over two sessions, and the logic gates skill was practiced over three sessions. In each task, participants were exposed to two rule sequences during training trials. For example, all problems in the logic gates required the solution of three logic gates in a sequence, and the sequence of rules was either *and-or-and* and *or-and-or*. For each of the two sequences, there were six distinct instances that differed in the pattern of digital input to the gates. On the final session of each skill task, blocks of transfer trials were introduced without warning following several training blocks. The transfer blocks contained trials with both the original training sequences and trials with new sequences that began like the old sequences but ended differently (*and-or-or* and *or-and-and*). As in previous experiments with different versions of these skill tasks, participants showed a substantial increase in errors on the new sequence trials compared to old. Figures 16 and 17 show the training and transfer trial data from the two skill tasks in this study.

The working memory measures were adapted from previous research on working memory by USAF Armstrong Laboratory's LAMP project. The Alphabet Working Memory task required participants to temporarily store 2 letters of the alphabet (e.g., T L) and recode them to two new letters either forward or backward in the alphabet (e.g., TL - 2 = RJ). The Continuous Opposites verbal working memory task required participants to remember the last three words from a continuous string of single syllable words (e.g., big, fast, high, hot, slow ....). When words appeared in yellow rather than white, the participant had to remember the opposite meaning of the stimulus. In previous research, these measures had high loadings on general working memory factors (e.g., Kyllonen & Christal, 1990).

There were four measures of attention processes. The intent was to measure the ability to disengage attention from expected cognitive processes, because this is posited to be the primary attention mechanism underlying error avoidance under varied task demands. Two tasks were loosely modeled after the Posner attention disengagement paradigm (Posner, Nissen, & Ogden, 1978). The primary difference was that these tasks required a disengagement of cognitive processes whereas Posner's paradigm required disengagement of perceptual processes. The Word Disengagement task presented word pairs for which participants had to decide whether they had similar or different meanings on 80% of the trials (these trials had a yellow frame surrounding the display). On the other 20% of the trials, the frame was blue and participants had to decide if the words had similar physical appearance (i.e., both lower case or both upper case versus different case). Across all trials, case similarity was crossed with meaning similarity such that
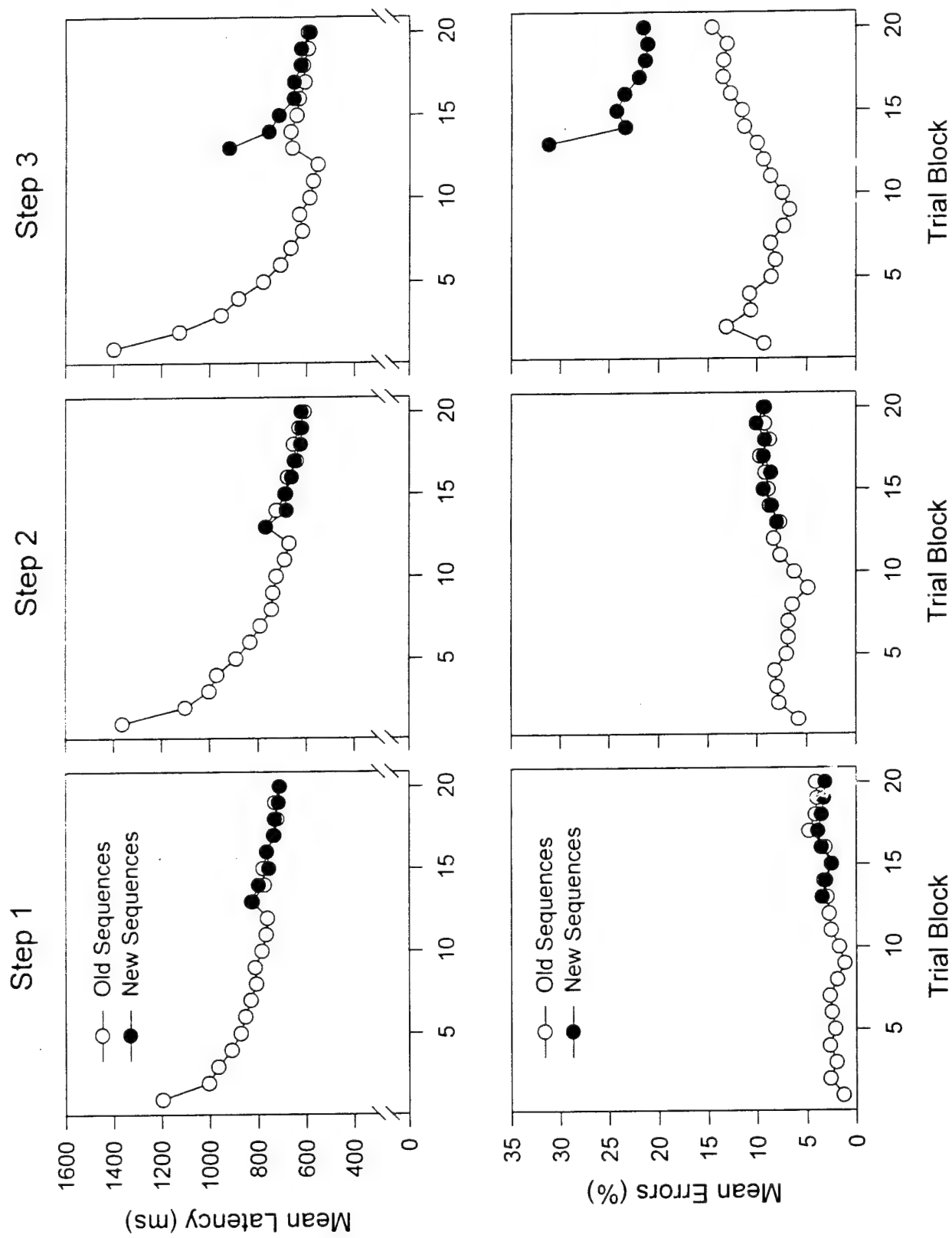
Figure 16: Mean Latency and Errors for Simple Number Reduction Training and Transfer Trials
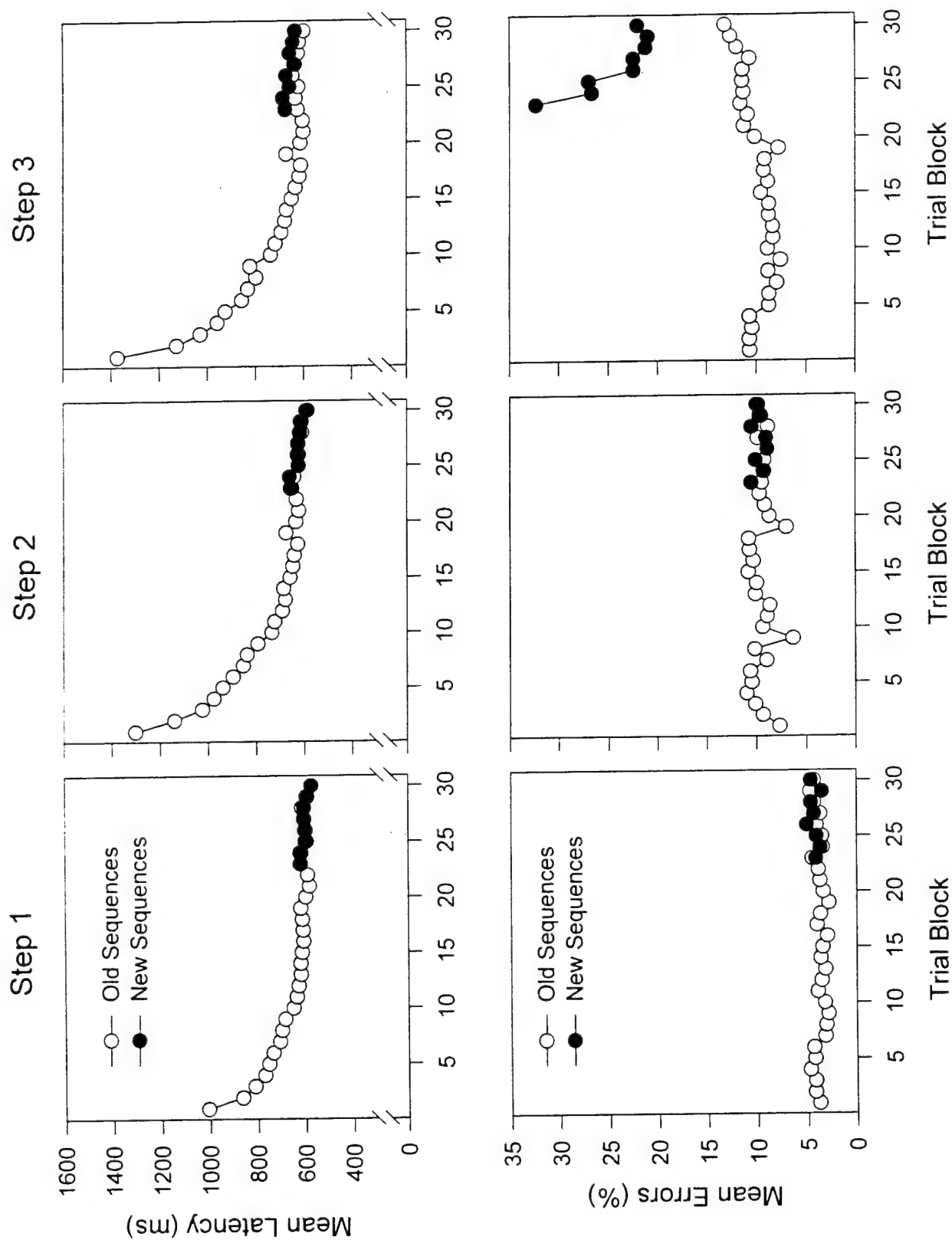
Figure 17: Mean Latency and Errors for Simple Logic Gates Training and Transfer Trials

there were equal numbers of the four possible combinations. Disengagement cost was assessed by latency and error increases associated with switching to case comparison trials (the 20%) relative to baseline latency and error rates when 100% of the trials were case comparisons. The second disengagement measure was similar except that the stimuli were single digits (1,2,3,4,6,7,8,9). These were displayed in a large format, with each digit equaling approximately 15 cm in height. The large format numbers (global numbers) were comprised of small characters (local numbers) that could be either consistent or inconsistent with the global numbers. On 80% of the trials signaled by green character color, participants had to evaluate whether the global number was less than or greater than 5. On the remaining 20% of the trials (signaled by blue character color), participants had to make the same size judgments about the local numbers. Again, disengagement cost was assessed by latency and error increases associated with switching to local size judgments (the 20%). The remaining two attention tasks were variations of the stroop task (MacLeod, 1991). A computerized version of the original color-word stroop task was given, and number stroop task was also developed following the work of Morton (1969).

Finally, there were two verbal knowledge tests and two perceptual speed tests. Vocabulary and general knowledge questions were taken from previous LAMP tests for these two knowledge constructs (Kyllonen, Woltz, Christal, Tirre, Shute, & Chaiken, 1990). The two perceptual speed tasks were modeled after the Finding A's and String Comparison tests of the Educational Testing Service (ETS) Kit of Cognitive Ability Measures (Ekstrom, French, Harman, & Dermen, 1976).

The most important finding from this study was that the attention measures uniquely predicted performance errors in near-transfer conditions that produce strong-but-wrong errors. Historically, it has been difficult to find ability measures other than some perceptual and psychomotor tests that predict skilled performance in the later stages of learning (see Ackerman, 1987, 1990; Fleishman, 1972). As shown in Figure 18, we identified three processing ability factors, attention (disengagement), working memory and speed, and general verbal knowledge. Although the knowledge and working memory/speed factors were positively related, the attention disengagement factor was unrelated to the other factors. Knowledge had no unique predictive relationship with the skill transfer errors. Working memory/Speed had a small but statistically significant relationship. Of primary importance, the attention disengagement factor was a strong predictor of skill transfer errors.

This evidence is important for two reasons. First, attention disengagement is a theoretically import cognitive construct that has previously not been thoroughly investigated. These data suggest that it may be important in some forms of near transfer. Second, most research on skill learning focuses on performance speed and ignores performance errors. Our research has attempted to understand error making during skilled performance. Although practice leads to fewer errors, simple undetected slips by otherwise skilled performers can be disastrous in critical jobs. Attaining high levels of skill in mental tasks brings certain inherent risks of some types of errors due to the fact that highly skilled performance is under less conscious control than earlier stages of skill. Understanding both the mechanisms underlying such errors and potential methods for reducing their likelihood could have important applications in many Air Force work environments.
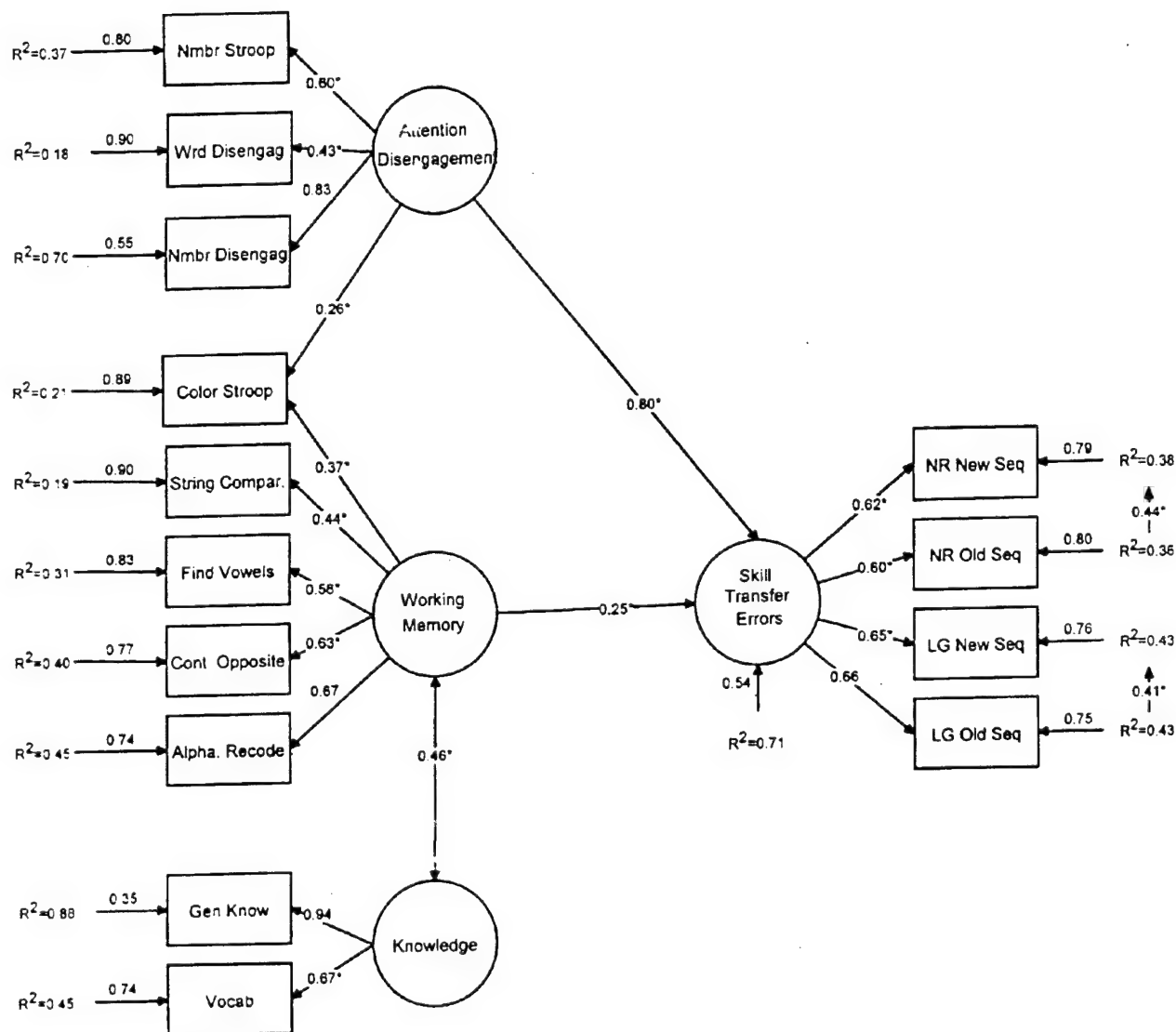
Model Fit Statistics: Chi Sq. = 102.80, p = .01; NNFI = 0.92; CFI = 0.94

Figure 18: Structural Model for the Prediction of Skilled Performance Errors from Attention Disengagement, Working Memory/Processing Speed, and Knowledge

Training Methods to Reduce Errors

We conducted one large-scale experiment to investigate the impact of two training variables on the likelihood of strong-but-wrong transfer errors. This study is reported in greater detail elsewhere (Gyll, 2000). The first training factor investigated was variability in processing sequences that learners experienced during training. Variability of experience during training has generally been linked to improved transfer in skills (Schmidt & Bjork, 1992). We investigated specifically whether variability in the processing sequences experiences improves transfer (i.e., reduces negative transfer).

The second training factor investigated the impact of attention flexibility training during skill acquisition on strong-but-wrong transfer errors. The attention flexibility training was based generally on the variable priority training method that has been shown to improve transfer in dual task experiments (Brickner & Gopher, 1981; Gopher, 1993; Gopher, Weil, & Seigel, 1989; Kramer, Larish, & Strayer, 1995). In the variable priority training method, participants are instructed during separate phases of training to place different degrees of emphasis on the different dual task components. The prior evidence suggests that skill practice under this training method produces better post-training performance and improved transfer to new dual tasks. Learners purportedly develop more flexible representations of the skill components. We adapted this approach in an attempt to develop greater flexibility during transfer to new processing sequences within a single skill. During skill training, we had participants alternate between emphasizing performance speed and performance accuracy. The intend was to develop greater flexibility in being able to speed up and slow down during skill performance, which should help during transfer to new sequences which require disengaging from fast, well-practiced operations and instead executing less familiar operations.

We conducted this research at USAF Armstrong Laboratory, using subjects and facilities at Lackland AFB, Texas. A total of 762 enlisted personnel participated during the final two weeks of their basic training. Participants were randomly assigned to one of eight different training conditions that presented a 3-step number reduction skill. Sample sizes ranged from 90 to 99 individuals in each of the eight conditions. All participants performed three phases of training and one phase of transfer. As in several previous studies, the new sequence trials introduced during transfer differed from training sequences only in the 3rd and final step. We varied three training factors between groups in a 2x2x2 design: (a) number of different sequences presented during training (2 vs. 4), (b) amount of practice in each training phase (4 vs. 8 trial blocks per phase), and (c) consistent versus variable speed/accuracy instructions. Those who received the variable speed/accuracy condition were instructed to respond quickly to stimuli presented in green, even if they made more errors, and to respond more slowly and carefully to stimuli presented in yellow. In the first two phases of training, green and yellow stimuli were alternated between trial blocks (odd blocks were green and even blocks were yellow). In the final training phase, color was varied randomly between trials within each block.

The mean error and latency data for training and transfer are presented in Figures 19–26. Figures 19-22 present the low practice conditions, and Figures 23-26 present the high practice conditions. The first two figures of each set (Figures 19-20 and Figures 23-24) present the consistent speed/accuracy conditions, and the last two figures of each set (Figures 21-22 and Figures 25-26) present the variable speed/accuracy conditions.

Finally, the odd numbered figures present conditions with only two training sequences and the even numbered figures present conditions with four training sequences.

Inspection of Figures 19-26 reveals the general finding consistent with previous evidence that the introduction of new processing sequences in the transfer phase (Blocks 25-30) produced errors in Step 3 of new sequences. Step 3 was the only step in which new sequences differed from old sequences. The questions of interest in this study were whether these errors were reduced by any of the training manipulations.

First, it can be readily seen in the figures that the number of training sequences had the greatest impact on strong-but-wrong transfer errors. The even numbered figures (4 training sequences) show substantially fewer transfer errors than the odd numbered figures (2 training sequences), $p < .05$. This indicates that the general principle of greater training variability improving transfer also applies to sequence memory and its role in negative transfer. Exposure to greater variability of processing sequences reduced the likelihood of negative transfer errors when new sequences were encountered.

Second, the amount of practice on training sequences had a smaller but reliable impact on transfer performance. Those with more practice during training were had shorter average response latency during transfer, but they made more new sequence errors. This is consistent with previous findings (Woltz, et. al, 2000), and it demonstrates that strong-but-wrong transfer errors are more likely among highly skilled individuals than among novices.

Given the impact of amount of training on transfer errors, it was conceivable that the effect of number of training sequences on errors simply reflected a practice effect rather than a variability effect. When fewer sequences were presented during training, there was more practice on each sequence (training block size was constant in all conditions), and this could have accounted for the increased number of errors. To test these competing hypotheses, we compared the transfer error rates of Low Practice 2-Sequence conditions (Figures 19-20) with High Practice 4-Sequence conditions (Figures 25-26). The amount of practice per training sequence was equivalent in these two sets of conditions. As can be seen in these figures, there were substantially more errors in the Low Practice 2-Sequence conditions. This suggests that variability in the representation of processing sequence knowledge had an important impact on reducing transfer errors beyond the impact of the strength of this knowledge.

Finally, the results indicated that the attention flexibility training did not have the desired effect of reducing transfer errors. The latency data during the three training phases indicated that participants were able to slow down and improve accuracy when instructed to do so both between blocks (Blocks 1-16) and between trials within blocks (Blocks 17-24). This training did improve transfer latency in that those in the variable speed/accuracy conditions were faster during transfer than those in the consistent speed/accuracy conditions ($p < .05$). However, this training manipulation did not affect transfer error rate ($p > .10$). Even when we selected only those participants who showed the greatest performance difference between the different instructional sets (i.e., those who complied most with training instructions), we found no effect of the training on transfer errors. Thus, this method of flexibility training appeared to be ineffective at reducing the likelihood of errors due to strong-but-wrong processing sequence execution.

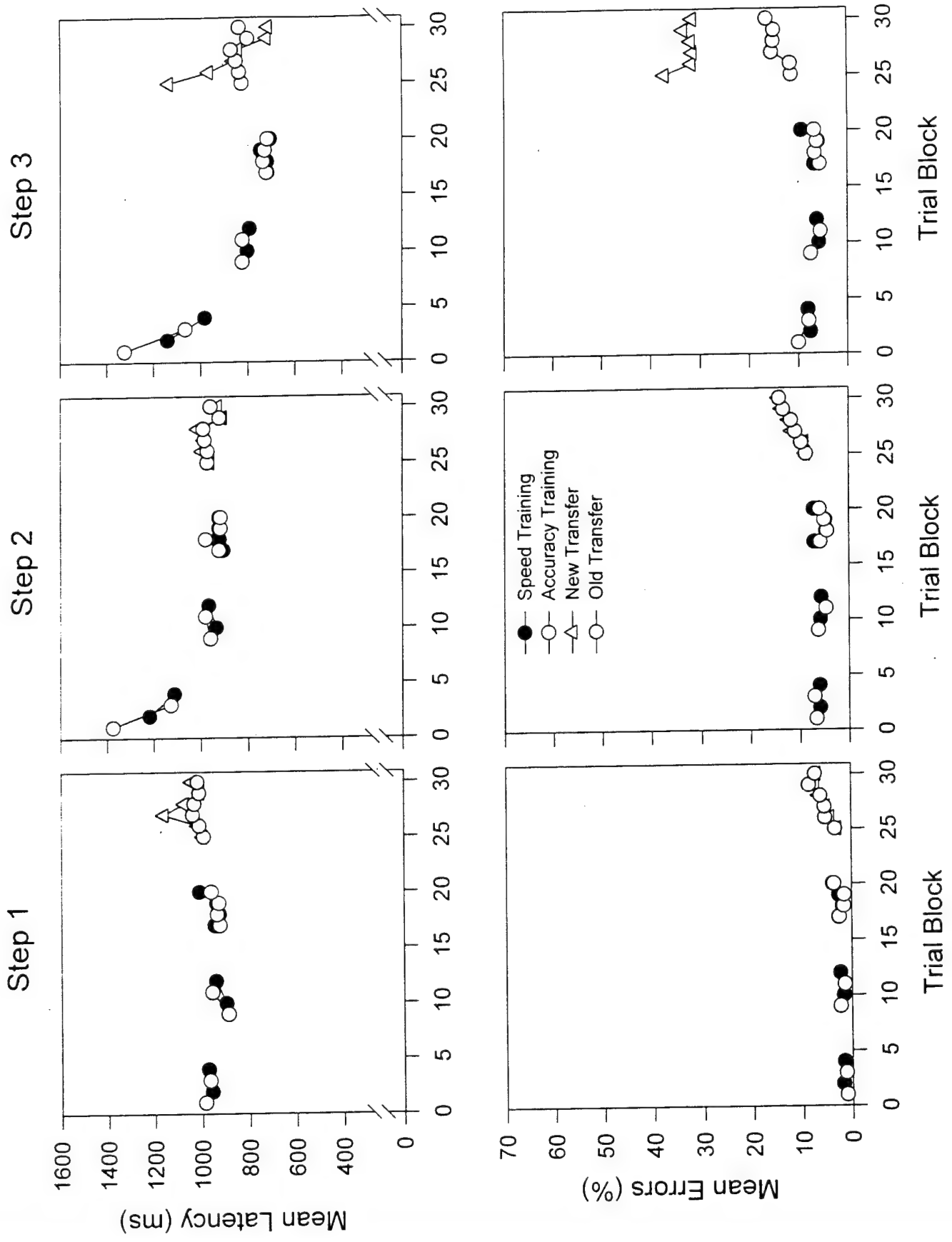Figure 19. Low Practice, 2 Training Sequences, Consistent Speed/Accuracy Training

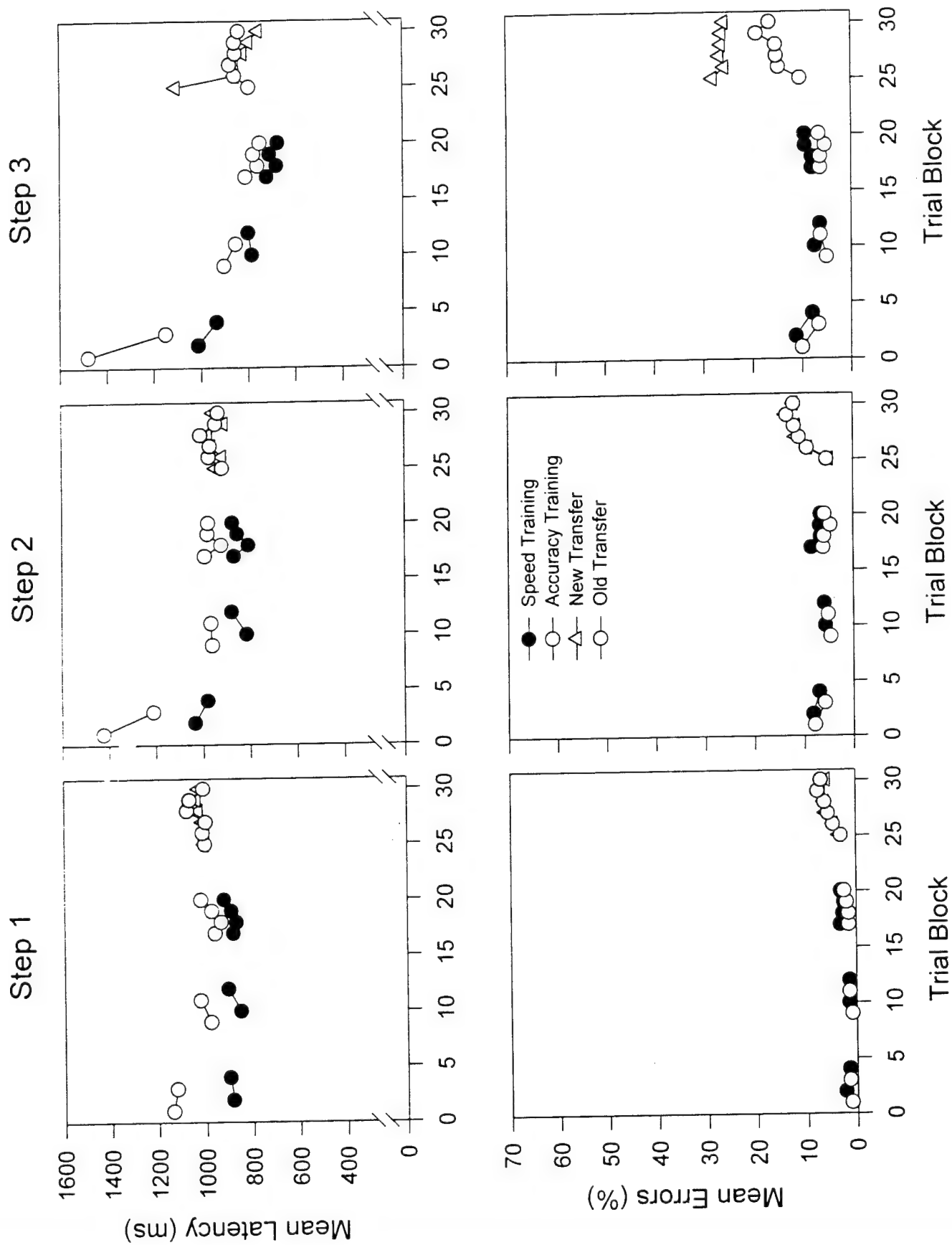Figure 20. Low Practice, 2 Training Sequences, Variable Speed/Accuracy Training

Figure 21. Low Practice, 4 Training Sequences, Consistent Speed/Accuracy Training
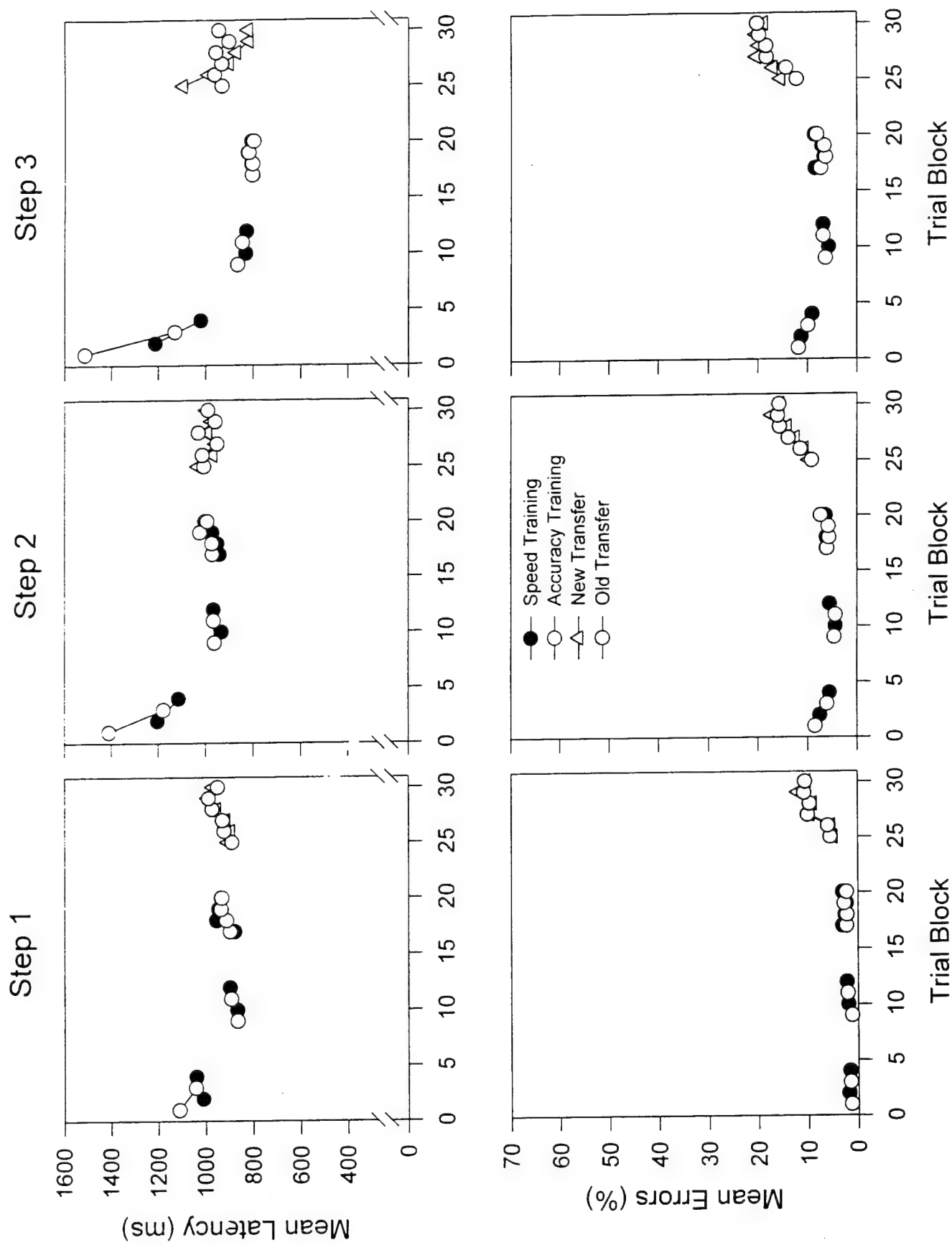
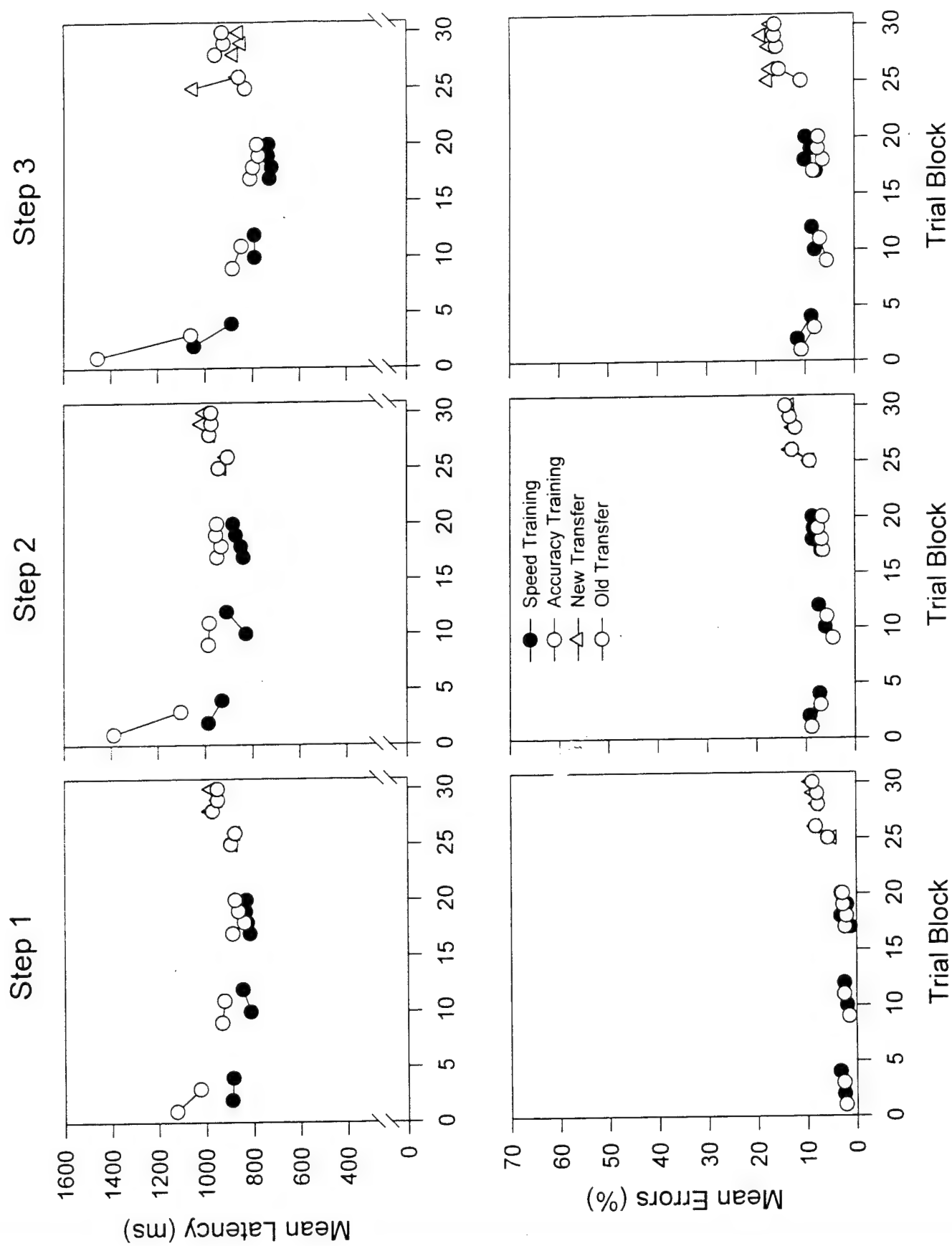Figure 22. Low Practice, 2 Training Sequences, Variable Speed Accuracy Training

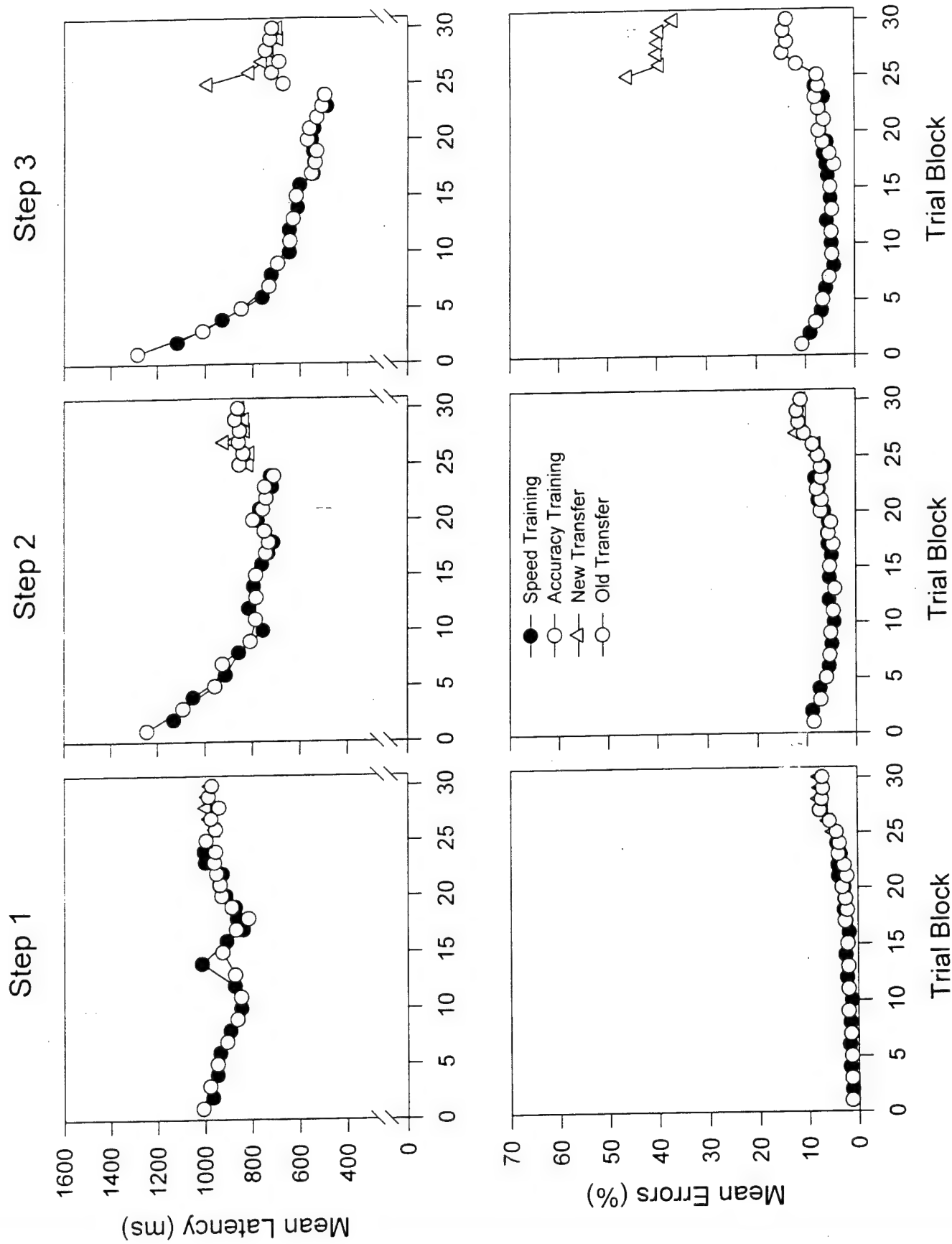Figure 23. High Practice, 2 Training Sequences, Consistent Speed/Accuracy Training

Figure 24. High Practice, 2 Sequences, Variable Speed/Accuracy Training
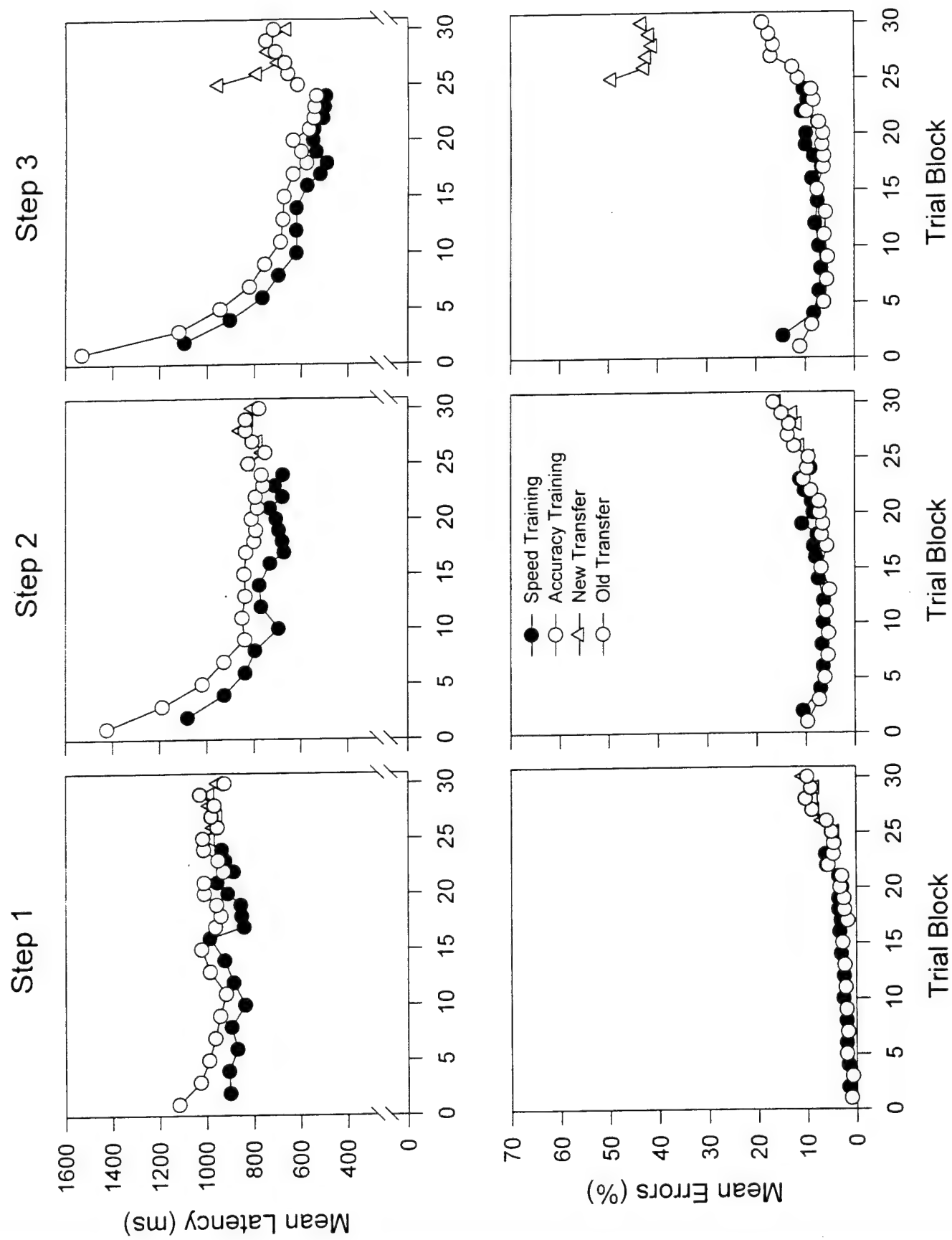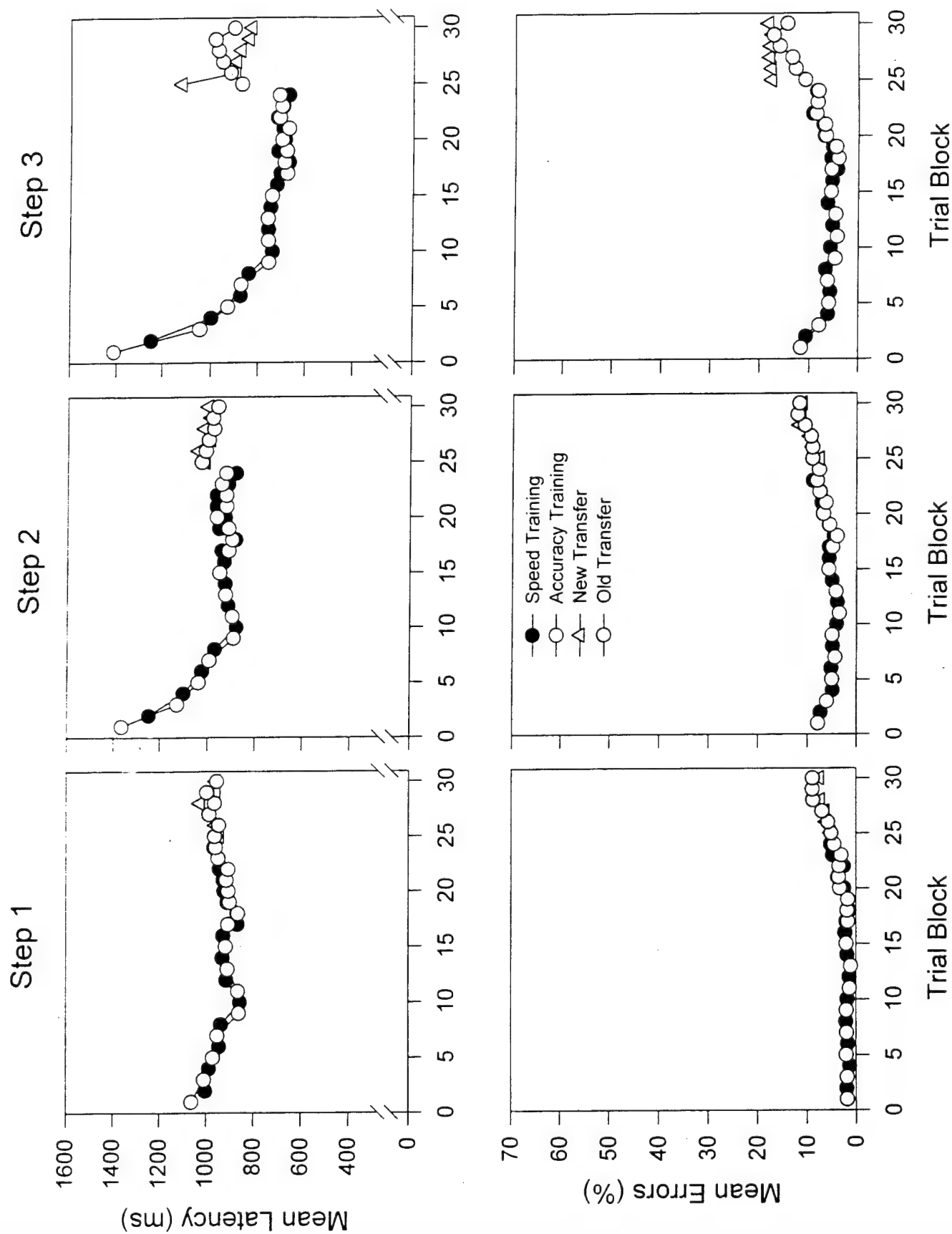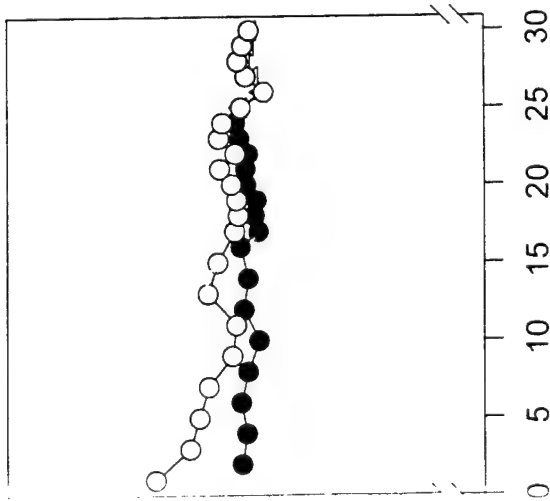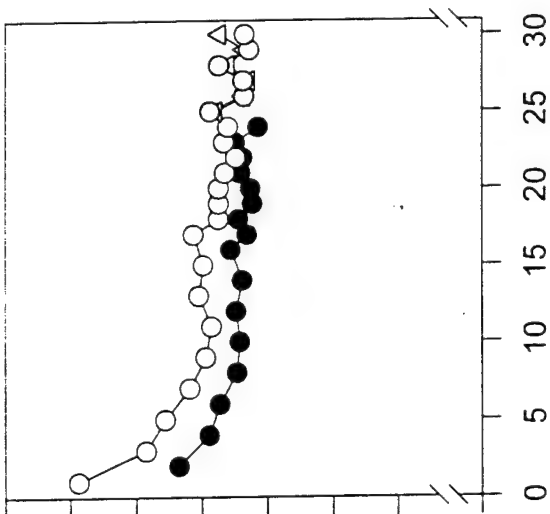
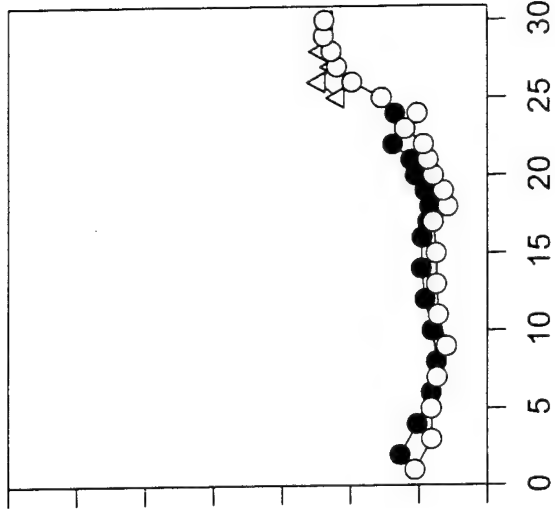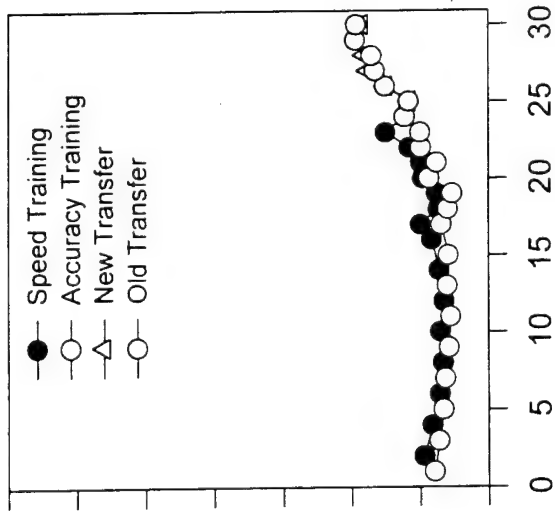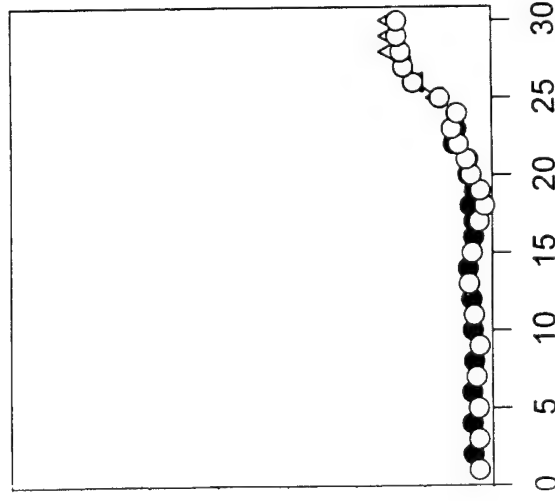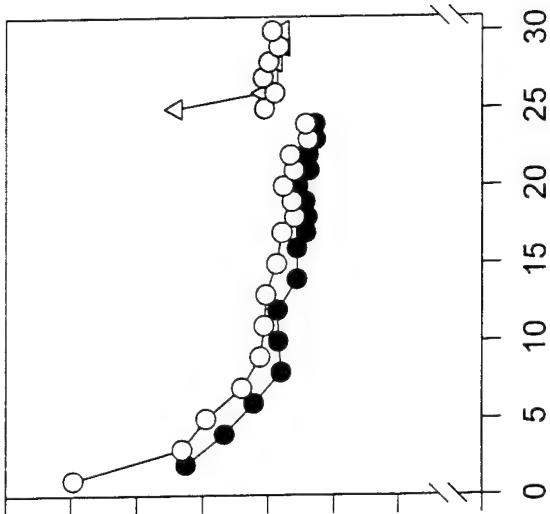Figure 25. High Practice, 4 Training Sequences, Consistent Speed/Accuracy Training

High Practice, 4 Training Sequences, Variable Speed/Accuracy Training

Step 1

Step 2

Step 3

Speed Training
Accuracy Training
New Transfer
Old Transfer

Trial Block

Summary of Conclusions and Implications

Under certain performance conditions, processing sequence knowledge can be misapplied and result in skill performance errors. Although we studied these performance conditions within controlled laboratory experiments, we believe the phenomena we studied in the laboratory are similar to those found in real world settings. First, many real world tasks have a sequential processing structure (e.g., diagnosis, computational problems, grammatical rule application, checklists, etc.). Second, we found evidence of strong-but-wrong errors due to sequence misapplication in every cascaded sequential processing skill that we investigated. We varied task content as well as task complexity and always found that we could produce these errors. Third, anecdotes of strong-but-wrong sequence errors are relatively common in everyday life (Norman, 1981; Reason, 1990). Thus, we believe strong-but-wrong sequence application errors represent an important class of errors that could be instrumental in critical operational environments (e.g., aviation, weapons control, medical diagnosis, etc.). Here we summarize the conclusions drawn from our research on these errors.

One finding of potential importance was that strong-but-wrong near transfer errors were more likely among highly skilled performers compared to less skill performers. This is in stark contrast to most other forms of performance errors (e.g., working memory errors during early stages of skill acquisition). This finding has theoretical significance in providing a relatively rare demonstration of negative transfer (see Woltz, et. al, 2000 for a discussion of this issue). In addition, it has some implications for understanding the complexity of this error type in some operational environments. In general, greater experience with any skill engenders confidence by the performer. Greater confidence often leads a performer to invest fewer attentional resources in monitoring performance for possible mistakes. This may be one reason why these errors are difficult to detect in the performer's awareness. They occur at later stages of skill acquisition when learners have automated many task components and probably feel quite confident in their performance ability. This makes this form of error particularly dangerous in environments where undetected slips could be disastrous.

Related to this problem, the sequence memory that underlies the strong-but-wrong errors appears to be implicit rather than explicit in nature. That is, skill performance is measurably affected by sequence memory, yet skilled performers appear to have little or no ability to consciously retrieve the processing sequence information. This knowledge appears to be available only in the context of skill performance, and even then it does not appear to be available to conscious introspection. Consequently, even if performers attempts to monitor performance during later stages of skill acquisition, they appear to lack introspective ability into the knowledge structures that would presumably help them understand the strong-but-wrong error threat.

Related to these issues, we found that strong-but-wrong errors occurring after extensive practice are generally undetected by the performers. The detected errors that occur during skill transfer appear to be primarily motor errors rather than cognitive errors. It appears as though error detection by skilled performers depends on the existence of an overt action that can be compared to a specific intention. In cascaded sequential processing, the result of cognitive actions typically can only be compared to general rather than specific intentions. That is, intentions in this case usually reflect a

goal of performing an operation, rather than arriving at a specific outcome. Again, the undetected nature of strong-but-wrong errors exacerbates the problem of cognitive mistakes in operational environments. Critical performance slips can only be corrected if they are detected by the performer prior to a catastrophic outcome.

Our evidence was consistent with the idea that sequence knowledge is represented in memory as linked associations that function by prior steps in a sequence priming subsequent steps. Our evidence ruled out chunking or composition as a form of sequence representation (i.e., all-or-none execution of entire sequences), and it also ruled out simple dyad representation of individual step transitions. This finding suggests that processing sequence knowledge can be quite complex in nature, but that there is at least some chance for attentional control within a sequence in that execution does not appear to be all or nothing. This suggests that different training strategies that encourage monitoring may be useful to investigate.

We investigated whether some individuals consistently make more strong-but-wrong errors than others, and if so, what cognitive characteristics describe these individuals. We found that individual error frequency during transfer generalized across skills. In previous research we found that self-reported propensity for cognitive slips and errors did not predict these strong-but-wrong skill performance errors. In the current effort, we also found that general cognitive aptitudes (working memory capacity, processing speed, general knowledge) that predict many forms of cognitive performance had little or no predictive ability for strong-but-wrong transfer errors. However, we did find a specific attention control process that had substantial ability to predict these skill transfer errors. We developed multiple, independent measures of the ability to disengage and re-engage attention, and we found these measures made unique and significant contributions in explaining sample variance in skill transfer errors. Furthermore, these measures did not correlated substantially with more traditional measures of cognitive ability. In total these findings suggested that the ability to avoid strong-but-wrong errors under performance conditions that promote these errors may depend on very specific attention processes. Again future research could explore in more detail the nature of this processing ability and whether it can be affected with practice or training.

Finally, we investigated several training factors that could influence the likelihood of transfer errors due to misapplication of sequence knowledge. As mentioned earlier, we found in several experiments that the amount of practice in a skill affects the likelihood of strong-but-wrong errors (more practice leads to more errors). We also found that the number of different processing sequences experienced during training affected the likelihood of transfer errors. Extensive exposure to a limited set of processing sequences makes transfer errors most likely. Paradoxically, it is these conditions that produce the most impressive rates of skill acquisition. Consequently, the lure in training environments may be to design skill practice using a limited range of exemplars because it leads to rapid skill acquisition. However, our evidence suggests that this could have negative consequences when more variable performance conditions are experienced in the operational environment.

There are several directions for future research that we believe could be important. First, we believe it would be important to understand the influence of fatigue and stress on the propensity of skilled performers to make errors. Highly trained performers of complex skills must often perform their jobs under less than ideal physical

and mental conditions. Understanding the degree to which fatigue and stress impact the likelihood of skilled performance errors could be of practice importance in military, industrial, and medical settings where such errors are very consequential. Second, we believe that the topic of skilled performance errors is ideal for furthering the theoretical understanding of the interplay of automatic memory-driven processes and attention-driven volitional processes. In our research we attempted to standardize our instructions and feedback to subjects and to minimize their awareness of the performance errors. However, we believe that it would be important in future research to investigate volitional processes that could reduce errors of this type, primarily by manipulating both instructions and feedback about errors. Finally, we have identified distinct memory components that underlie the acquisition of sequential processing skills (e.g., declarative memory for task rules, procedural memory for component operations, procedural memory for processing instances, and procedural memory for processing sequences). For both theoretical and practical reasons, we believe it would be important in future research to investigate the forgetting rates associated with each of these components. This could be of benefit to the applied problem of refreshing and retraining skills that were previously learned.

--

References

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. Psychological Bulletin, 102(1), 3-27.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. Journal of Experimental Psychology General, 117(3), 288-318.

Ackerman, P. L. (1990). A correlational analysis of skill specificity: Learning, abilities, and individual differences. Journal of Experimental Psychology Learning, Memory, and Cognition, 16(5), 883-901.

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. Journal of Applied Psychology, 77(5), 598-614.

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. Psychological Review, 94, 192-210.

Anderson, J. R. (1989). The analogical origins of errors in problem solving. In D. Klahr & K. Kotovsky (Eds.), Information processing: The impact of Herbert A. Simon (pp. 343-371). Hillsdale, NJ: Erlbaum

Anderson, J. R. (1993). Rules of the mind. Hillsdale, NJ: Erlbaum.

Bell, B.G., Gardner, M.K., & Woltz, D.J. (1997). Individual differences in skilled performance. Learning and Individual Differences, 9(1), 43-61.

Blessing, S. B., & Anderson, J. R. (1996). How people learn to skip steps. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(3), 576.

Bogner, M. S. (1994). Human error in medicine. Hillsdale, NJ: Erlbaum.

Broadbent, D. E., Baddeley, A. D., & Reason, J. T. (1990). Human factors in hazardous situations. Oxford, UK: Oxford University Press.

Carlson, R. A., & Lundy, D. H. (1992). Consistency and restructuring in learning cognitive procedural sequences. Journal of Experimental Psychology Learning, Memory, and Cognition, 18(1), 127-141.

Carlson, R. A., & Schneider, W. (1989). Practice effects and composition: A reply to Anderson. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 531-533.

Carlson, R. A., & Shin, J. C. (1996). Practice schedules and subgoal instantiation in cascaded problem solving. Journal of Experimental Psychology Learning, Memory, and Cognition, 22(1), 157-168.

Carlson, R. A., Sullivan, M. A., & Schneider, W. (1989). Practice and working memory effects in building procedural skill. Journal of Experimental Psychology Learning, Memory, and Cognition, 15(3), 517-526.

Carlson, R. A., & Yaure, R. G. (1990). Practice schedules and the use of component skills in problem solving. Journal of Experimental Psychology Learning, Memory, and Cognition, 16(3), 484-496.

Charness, N., & Campbell, J. I. (1988). Acquiring skill at mental calculation in adulthood: A task decomposition. Journal of Experimental Psychology: General, 117(2), 115-129.

Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16(1), 17-30.

Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill. Ergonomics, 2, 153-166.

Curran, T., & Keele, S. W. (1993). Attentional and nonattentional forms of sequence learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19(1), 189-202.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). Kit of factor-referenced cognitive tests. Princeton, NJ: Educational Testing Service.

Elio, R. (1986). Representation of similar well-learned cognitive procedures. Cognitive Science, 10, 41-73.

Fendrich, D. W., Healy, A. F., & Bourne, L. E. (1991). Long-term repetition effects for motoric and perceptual procedures. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17(1), 137-151.

Fitts, P. M., & Posner, M. I. (1967). Human Performance. Belmont, CA: Brooks Cole.

Fleishman, E. A. (1972). On the relation between abilities, learning, and human performance. American Psychologist, 27, 1017-1032.

Frensch, P. A. (1991). Transfer of composed knowledge in a multistep serial task. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 997-1016.

Frensch, P. A. (1994). Composition during serial learning: A serial position effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(2), 423-442.

Gopher, D. (1993). The skill of attention control: Acquisition and execution of attention strategies. In D. Meyer & S. Kornblum (Eds.), Attention and Performance XIV . Hillsdale, NJ: Erlbaum.

Gopher, D., Weil, M., & Siegel, D. (1989). Practice under changing priorities: An approach to training of complex skills. Acta Psychologica, 71, 147-179.

Gyll, S.P. (2000). Cognitive flexibility training in the reduction of skilled performance errors. Unpublished doctoral dissertation, University of Utah, Salt Lake City.

Hayes, W. L. (1988). Statistics. New York: Hold, Rinehart, and Winston.

Heckhausen, H., & Beckman, J. (1990). Intentional action and action slips. Psychological Review, 97, 36-48.

Johnston, W. A., Hawley, K. J., & Elliott, J. M. (1991). Contribution of perceptual fluency to recognition judgments. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 210-223.

Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. Journal of Experimental Psychology Learning, Memory, and Cognition, 22(1), 169-181.

Kramer, A. F., Larish, J. F., & Strayer, D. L. (1995). Training for attentional control in dual task settings: A comparison of young and old adults. Journal of Experimental Psychology: Applied, 1(1), 50-76.

Kyllonen, P. C. (1995). The cognitive abilities measurement (CAM) battery version 4, reference manual (Technical Paper AL/HR-TP-0000). Brooks AFB: Armstrong Laboratory.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?~. Intelligence, 14(4), 389-433.

Kyllonen, P. C., & Stephens, D. L. (1990). Cognitive abilities as determinants of success in acquiring logic skill. Special Issue: Frontiers of learning and individual differences: Paradigms in transition. Learning and Individual Differences, 2(2), 129-160.

LaBerge, D. (1973). Attention and the measurement of perceptual learning. Memory and Cognition, 1, 268-276.

Larkin, A. A., Woltz, D. J., Reynolds, R. E., & Clark, E. (1996). Conceptual priming differences and reading ability. Contemporary Educational Psychology, 21(3), 279-303.

Lewicki, P., Czyzewska, M., & Hoffman, H. (1987). Unconscious acquisition of complex procedural knowledge. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13(4), 523-530.

Lewicki, P., Hill, T., & Bizot, E. (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. Cognitive Psychology, 20(1), 24-37.

Logan, G. D. (1988). Toward an instance theory of automatization. Psychological Review, 95(4), 492-527.

Luchins, A. S. (1942). Mechanization in problem solving. Psychological Monographs, 54, (Whole No. 248).

Lundy, D. H., Wenger, J. L., Schmidt, R. J., & Carlson, R. A. (1994). Serial step learning of cognitive sequences. Journal of Experimental Psychology Learning, Memory, and Cognition, 20(5), 1183-1195.

MacKay, D. G. (1982). The problems of flexibility fluency, and speed-accuracy trade-off in skilled behavior. Psychological Review, 89, 483-506.

MacKay, D. G. (1987). The organization of perception and action. New York: Springer-Verlag.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review.

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, R. J., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15(6), 1083-1100.

Morton, J. (1969). Categories of interference: Verbal mediation and conflict in card sorting. British Journal of Psychology, 60, 329-346.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), Cognitive skills and their acquisition (pp. 1-55). Hillsdale, NJ: Erlbaum.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. Cognitive Psychology, 19(1), 1-32.

Nissen, M. J., Willingham, D., & Hartman, M. (1989). Explicit and implicit remembering: When is learning preserved in amnesia? Neuropsychologia, 27(3), 341-352.

Norman, D. A. (1981). Categorization of action slips. Psychological Review, 88(1), 1-15.
    Rasmussen, J., Duncan, K., & Leplat, J. (Eds.). (1987). New technologies and
    human error. New York: Wiley.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of
    signals. Journal of Experimental Psychology: General, 109, 160-174.

Reason, J. T. (1990). Human Error. Cambridge, England: Cambridge University Press.

Rosenbloom, P. S., Laird, J. E., & Newell, A. (1989). The chunking of skill and
    knowledge. In B. A. G. Elsendoorn & H. Bouma (Eds.), Working models of
    human perception (pp. 391-410). London, England UK: Academic Press, Inc.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information
    processsing. I. Detection, search, and attention. Psychological Review, 84, 1-66.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common
    principles in three paradigms suggest new concepts for training. Psychological
    Science, 3(4), 207-217.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information
    processing: II. Perceptual learning, automatic attending, and a general theory.
    Psychological Review, 84, 127-190.

Singley, M. K., & Anderson, J. R. (1989). The transfer of cognitive skill. Cambridge,
    MA: Harvard University Press.

Stadler, M. A. (1989). On learning complex procedural knowledge. Journal of
    Experimental Psychology: Learning, Memory, and Cognition, 15(6), 1061-1069.

Thurstone, L. L., & Thurstone, T. G. (1941). Factorial studies of intelligence. Chicago:
    University of Chicago Press.

Wenger, J. L., & Carlson, R. A. (1996). Cognitive sequence knowledge: What is learned?
    Journal of Experimental Psychology Learning, Memory, and Cognition, 22(3),
    599-619.

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of
    procedural knowledge. Journal of Experimental Psychology: Learning, Memory,
    and Cognition, 15(6), 1047-1060.

Woltz, D. J. (1988). An investigation of the role of working memory in procedural skill
    acquisition. Journal of Experimental Psychology: General, 117, 319-331.

Woltz, D. J. (1999). Individual differences in priming: The roles of implicit facilitation
    from prior processing. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.),
    Learning and individual differences: Process, trait, and content determinants (pp.
    135-156). Washington, DC: American Psychological Association.

Woltz, D. J., Bell, B. G., Kyllonen, P. C., & Gardner, M. K. (1996). Memory for order of
    operations in the acquisition and transfer of sequential cognitive skills. Journal of
    Experimental Psychology Learning, Memory, and Cognition, 22(2), 438-457.

Woltz, D. J., Gardner, M. K., & Bell, B.G. (2000). Negative transfer in sequential
    cognitive skills: Strong-but-wrong sequence application. Journal of Experimental
    Psychology Learning, Memory, and Cognition, 26, 601-625.

Woltz, D. J., Gardner, M. K., Bell, B. G., & Gyll, S. (in preparation). Negative transfer
    errors in logic gates skill acquisition.

Woltz, D. J., Gardner, M. K., & Gyll, S. (submitted for publication). The role of
    attention processes in near transfer of cognitive skills.

Appendix A

Publications and Submissions

Bell, B. G., Gardner, M. K., & Woltz, D. J. (1997). Individual differences in undetected errors in skilled cognitive performance. *Learning and Individual Differences, 9*, 43-61.

Gardner, M. K., Strayer, D. L., Woltz, D. J., & Hill, R. D. (2000). Cognitive skill acquisition, maintenance, and transfer in the elderly. In R. D. Hill, L. Backman, & A. S. Neely (Eds.), *Cognitive rehabilitation in old age*. Oxford, UK: Oxford University Press.

Gardner, M.K., Woltz, D.J., & Bell, B.G. (accepted for publication). Representation of memory for order of mental operations in cognitive tasks. *American Journal of Psychology*.

Woltz, D. J. (1999). Individual Differences in Priming: The Roles of Implicit Facilitation from Prior Processing. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Processes, traits, and content determinants*. Washington, DC: American Psychological Association.

Woltz, D. J. (under revision). *Direct priming of conceptually-driven mental operations*. Manuscript under revision for *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Woltz, D. J., Gardner, M. K., & Bell, B. G. (2000). Negative transfer in sequential cognitive skills: Strong-but-wrong sequence application. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 601-625.

Woltz, D. J., Gardner, M. K., & Gyll, S.P. (under review). *The role of attention processes in the near transfer of cognitive skills*. Manuscript under review by *Learning and Individual Differences*.

Appendix B
Participating Professionals

Dan J. Woltz, Ph.D., Associate Professor, Co-Principal Investigator
Michael K. Gardner, Ph.D., Professor, Co-Principal Investigator
John Kircher, Ph.D, Professor, Statistical Consultant
Brian G. Bell, Graduate Student (Ph.D) Research Assistant
       Dissertation Title: The effects of anziety on skilled performance errors
Sean Gyll, Graduate Student (Ph.D) Research Assistant
       Dissertation Title: Cognitive flexibility training in the reduction of skilled performance errors
Mary Hales, Graduate Student (Ph.D) Research Assistant
Angelique Scharine, Graduate Student (M.S.) Research Assistant
Scott Schliebner, Graduate Student (M.S.) Research Assistant
Alison Guthrie, Undergraduate Student (B.S.) Research Assistant
Yasamin Houseini, Undergraduate Student (B.S.) Research Assistant
Pon Insixiengmy, Undergraduate Student (B.S.) Research Assistant

Appendix C

New Discoveries, Inventions, and Patents

None

Footnotes

---

[1] In the simple version of number reduction, all responses were 1, 2 or 3. Participants could rest three fingers on these adjacent keys, and respond with intermediate answers without looking at the keyboard. However, the complex version had nine response alternatives, and participants would need to search the keyboard for each intermediate response. We believed that this would disrupt the acquisition of fluent transitions between component rules, so we required only a final response.

[ii] Several participants had missing data for one of the cells (e.g., they never called an old trial new). So, in this analysis there were 29 participants in the instance recognition condition and 24 participants in the sequence recognition condition.

[iii] The residual score method for indexing facilitation has generally yielded more reliable measures than other methods (Larkin, Woltz, Reynolds, & Clark, 1996; Woltz, 1999). This was also true in the current experiment. However, the internal consistency reliability estimate of the residual scores in the instance recognition group was only $\alpha = .28$. For the sequence recognition group, the internal consistency reliability was better at $\alpha = .65$.

[iv] The correlation for the instance group is somewhat difficult to interpret because the two variables had low reliability. As noted previously, the performance facilitation variable had an internal consistency estimate of .29. The d' measure for this group also had low internal consistency at $\alpha = .27$. These low reliabilities reflect a lack of consistent individual differences which could be due to the low magnitude of facilitation and a substantial reliance on guessing by most individuals during recognition. For the sequence group, the reliability estimates were higher ($\alpha = .65$ for performance facilitation and $\alpha = .74$ for d'). The greater reliability reflects the fact that some individuals were consistently better at recognizing old trials and some individuals were faster at performing old trials relative to new trials. However, these were not the same individuals, as indicated by the low correlation between facilitation and accuracy. Even when the correlation (-.12) was disattenuated for measurement error, it remained low (-.17).